Re-Thinking the Extended Mind

Moving Beyond the Machinery

John O'Regan

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of Doctor of Philosophy.

April 2010

Abstract

Proponents of the Extended Mind Thesis (EMT) argue that the mind literally extends into the world because mental states literally extend into the world. But the arguments presented in favour of these claims are compatible with a much weaker conclusion, expressed as the Extended Machinery of Mind Thesis (EMMT) that secures only the extension of the enablers of mental states.

What is required is a mark of the mental that can settle the constitutive versus enabling issue. Both sides of the debate accept non-derived content as a necessary condition on a state's being mental but this cannot settle the constitution versus enabling issue, meaning the debate has stagnated because there are no decisive moves left to make.

Thus, the strongest move for the EM theorist to make is to reject non-derived content as the mark of the mental and seek an alternative. Because enactivism rejects the representational view of mind then *if* it can be made to work as an account of mentality it offers promise with regard to the formation of a new mark of the mental on which a genuinely interesting EMT can be based.

Table of Contents

Acknowledgements **iv**

Introduction **1**

Chapter 1 **6**

Introduction 6
EMT vs. Content Externalism 6
Epistemic vs. Pragmatic Actions 8
Tetris 12
The Parity Principle 17
Otto 25
Mental States 32
Constitution vs. Enabling 41

Chapter 2 **46**

Introduction 46
EMT vs. Content Externalism 46
The Ontological Claim 55
The Manipulation Thesis 58
The Barking Dog & Evolutionary Support for EMT 60
Cognitive Extension 66
Vehicle Externalism 74
Constitution vs. Enabling 77

Chapter 3 **80**

Introduction 80
EMT vs. Content Externalism 80
Vehicles 83
The Unity of Consciousness 86
The Vehicle/Content Distinction 90
Extended Vehicles & the Dynamic Singularity 94
General Duplication 100

Colour Inversion & Left-Right Reversal 105
El Greco Worlds 113
Extended Vehicles 123
Constitution vs. Enabling 125

Chapter 4 **128**
Introduction 128
Not all Cognition is Extended 128
Differences 132
The Coupling/Constitution Fallacy 139
Cognitive Bloat 145
The Mark of the Mental 150

Chapter 5 **161**
Introduction 161
Functionalism & EMT 163
Functionalism & Mental States 175
EMT & Mental States 183
EMT & Cognitive Processes 193
Conclusion 194

Chapter 6 **196**
Introduction 196
The Mark of the Mental Re-Visited 197
Extended Representations 203
Enactivism 207
Enactive Perception 211
A New Mark of the Mental? 218
Conclusion 220

Conclusion **223**
Issues for Enactivism 227

Bibliography **229**

Acknowledgements

My thanks, firstly, to Mark Rowlands who got me interested in the Extended Mind and for supervising me through the early stages, without whom I wouldn't have started this project. Thanks also to the NUI whose generous scholarship allowed me to undertake my move to the UK. Thanks to Bruce Christianson for his ability to see the big picture and to Sam Coleman for his attention to detail and thanks to both for their encouragement and support.

My principal supervisor, Dan Hutto, went above and beyond the call of duty for a supervisor and for that I am immensely grateful, without him I would not have finished this project.

Thanks to Gillian for her support during the project and for putting up with me for the duration. My greatest thanks, finally, to my parents, Patrick and Rosarie O'Regan, for their patience, support and amazing generosity. I asked for more than I was entitled to and they offered more than I had a right to expect. This work, such as it is, is dedicated to them.

Introduction

The Extended Mind Thesis (EMT) has received a great deal of attention since Clark and Chalmers ignited the debate with the publication of 'The Extended Mind' in 1998, triggering a large volume of work from both critics and sympathisers (e.g. Rupert 2004; Adams & Aizawa 2001, 2005, 2008; Brook 2006, Sterelny 2004, Shapiro 2008, Clark 2001, 2003, 2007, 2008, 2009, Rowlands 1999, 2002, 2003, 2006, 2009; Menary 2006, 2007; Hurley 1998, 2000, Wheeler 2005) and at least three dedicated conferences.[1] This seminal paper can currently be found in readers for undergraduate philosophy courses (see Lycan & Prinz (eds.) *Mind and Cognition: An Anthology, 3rd edition*), illustrating that EMT has, in the relatively short time since its inception, established itself in the philosophical firmament. Its initial success notwithstanding, I argue that we must re-think the claims and commitments of EMT as currently formulated if a genuinely interesting argument for the extension of mind is to be developed.

I offer much needed analysis and clarification of the claims of EMT as presented by Clark and Chalmers. Doing so will reveal its core claims to be problematic because it accepts too much from traditional views of mind that it challenges on the issue of internalism. This is not just a descriptive analysis of the field. The debate has stagnated recently and my original analysis reveals that acceptance, by both sides, of the representationalist view of mind and, in particular, of non-derived content as a necessary condition on a state's being mental means there seems to be no decisive moves left to make.

[1] The Extended Mind: The Very Idea, The University of Hertfordshire 29th June – 1st July 2001, The Extended Mind II: Just when you thought it was safe to go back in the head, The University of Hertfordshire 10th -12th July 2006, and The Extended Mind Thesis in Theory and Applications, ZiF Bielefeld, Germany 23rd -25th November 2009.

The core claim to be defended in this work, then, is that when properly understood EMT does not yield a decisive argument for extension of the defining aspects of the mental. At best—although those aspects of mental states that are definitive of mentality do not extend—it is possible that the *functional aspects* of non-biologically basic mental states sometimes extend, in certain special cases. The principal conclusion of this thesis is that an interesting version of EMT has yet to be formulated: my aim is to set the stage for and motivate that development, and to say something about what it would require.

In order to establish the core thesis, I begin in chapter one with an explication of EMT as originally formulated by Clark and Chalmers and also examine more recent adjustments to this formulation by Clark. I suggest that Clark's exact position can be hard to pin down and argue that this is symptomatic of a tension in his work between what is advertised, which is nothing less that the literal extension of the mind into the world, and what the arguments can conclusively demonstrate, which is a more modest, and more easily established claim concerning the extension of the enablers of certain mental states. I also draw attention to the key distinction between cognitive processes and mental states. Clark and Chalmers point out that the extension of cognitive processes may not be sufficient to secure EMT proper since the extension of mind requires moving beyond mere processing. Thus, EMT requires the extension of mental states into the world.

Chapters two and three continue the expository work of the first chapter, focusing on the work of Mark Rowlands and Susan Hurley. Both authors are commonly taken to advocate a version of EMT but if this requires the extension of mental states then we can question whether or not they argue for a genuine EMT, since their focus is on the extension of cognitive processes and the vehicles of the unity of consciousness, respectively.

Chapter four describes some responses and objections that have been levelled at EMT in the literature and demonstrates that what is taken to be at stake in this debate, on both sides, is nothing less than the literal extension of the mental or the cognitive into the environment. I argue that each of the objections considered reduce to one key issue: constitution versus enabling. In order to settle questions concerning the location of the constituents of mental states as distinct from the mere enablers of mental states, we require an independently motivated mark of the mental. Clark argues that he can accept a proposal by Adams and Aizawa for non-derived content being the mark of the mental and maintain his argument for the extension of mind into the world so long as the extended aspects of mental states maintain a *special relationship* with internal non-derived contents.

In order to assess the possibility for success of this strategy I return in chapter five to re-examine the fundamental functionalist account of mental states and their properties that is presupposed by EMT in order to clarify precisely the nature of its claims. If functionalism alone were enough for developing an understanding of the mind then perhaps a broad functionalism of the sort espoused by Jackson and Pettit (1988) could yield an interesting (non-trivial) version of EMT. However, many hold that functionalism alone is not sufficient for understanding the mind. Consequently, it is thought to require supplementation with a representational theory of mind and, moreover, that it is the representational content of mental states that is definitive of mentality. But if the representational contents that individuate mental states are both definitive of mentality and internal then in what sense can the mental state be said to extend?

This analysis will reveal that, because of its exclusive focus on processes that *support* or *enable* mental states (e.g. belief-forming mechanisms)

proponents of EMT only give decisive arguments for a weaker thesis about the extended 'machinery' of the mind – not of mind itself, or of the mental. Understood in this way, EMT converts to EMMT (The Extended Machinery of Mind Thesis) and poses no interesting threat to internalism regarding mental states.

The investigation will expose that participants in the standard debates about EMT have been talking at cross purposes – i.e. that they have not clearly articulated the true implications of the proposal of EMT as set forth by Clark and Chalmers. Despite the enormous attention that Clark and Chalmers' proposals have attracted in recent years, if EMT is understood as EMMT it becomes clear that the thesis is of limited interest and that EMT has been falsely advertised.

Without arguments to secure the extension of the defining aspects of mental states into the world the only other possible move for defenders of EMT to make would be to argue that although not all aspects of mind extend and not even those that are definitive of mentality, it is at least possible that the *functional aspects* of non-biologically basic mental states sometimes extend, in certain special cases. But even if such a claim could be upheld it too undercuts the value and potential interest of EMT.

In the light of this analysis Chapter six considers the possibility of what an interesting version of EMT might look like and what it would require. I illustrate how Clark's unwillingness to engage the issue of the mark of the mental and propose an alternative to the representational view of mind stymies his capacity to develop a truly challenging version of EMT. But new theories of cognition are developing which deny that mental representations are the basis of mentality (see enactivism; Noë, Di Paolo, Thompson, O'Regan, Hutto) and suggest the possibility of a different, content-free 'mark of the mental'. *If* these approaches can be made to

work then an interesting version of EMT would be possible. I conclude with a brief summing up of the principal argument and lay out some possible questions for future research.

Chapter 1

Introduction

The primary purpose of this chapter is exposition—I explore and explain EMT as defended by Andy Clark and David Chalmers. I do this by considering arguments for thinking that playing Tetris is a case of extended cognition and for the attribution of extended beliefs to an Alzheimer's sufferer who uses a notebook as an 'external memory'. In the process I also highlight a major worry that will form the backbone of my thesis: there is a tension between how EMT is advertised and what can conclusively be argued for. As advertised, EMT is about the extension of mind into the world. But in this chapter I will demonstrate that Clark and Chalmers do not give us enough to secure this claim. All that is conclusively demonstrated is the extension of cognition-enabling and belief-forming mechanisms. The arguments for EMT presented by Clark and Chalmers are too weak, since they do not secure the extension of mental states into the world.

EMT vs. Content Externalism

EMT, in its various different incarnations is presented as introducing a form of externalism that is distinct from the more familiar Content Externalism of Putnam and Burge, and Clark and Chalmers introduce 'The Extended Mind' (1998) by drawing attention to this distinction:

> Where does the mind stop and the rest of the world begin? The question invites two standard replies. Some accept the demarcations of skin and skull, and say that what is outside the body is outside the mind. Others are impressed by arguments suggesting that the meaning of our words "just ain't in the head", and hold that this externalism about meaning carries over into an externalism about mind. We propose to pursue a

> third position. We advocate a very different sort of externalism: an *active externalism*, based on the active role of the environment in driving cognitive processes (Ibid: 7)

It is not entirely clear what Clark and Chalmers mean by the term 'driving' here. Indeed, it seems an odd choice of word if their view is that the environment can *constitute* part of the cognitive process (see below) and not merely be an add-on to, or enabler of, cognition. This sort of expression is key in highlighting the tension inherent in the claims of EMT.

But we can put these worries to one side for the moment and consider how EMT is presented as furthering the externalist agenda by challenging internalism in new and possibly more interesting ways when compared with Content Externalism. Roughly, the EM theorist argues not merely for the partial external *individuation* of mental state *contents* but the partial external *constitution* of mental states and cognitive processes. Clark and Chalmers argue that where content externalism invokes passive and distal external features that are situated at the end of a long causal chain to individuate and attribute mental content, their radical new externalism recruits active and proximal external features that are required for the production of behaviour. Because of the passive and distal nature of the external resources availed of by content externalism, they are not thought to play a causal or explanatory role in the here-and-now generation of action:

> In counterfactual cases where internal structure is held constant but external features are changed, behaviour looks just the same; so internal structure seems to be doing the crucial work. (Ibid: 9)

So, for example, if I and my twin-earth doppelganger are both thinking *that water is wet* then any resulting actions will be physically indistinguishable in spite of the fact that he is surrounded by XYZ and I am surrounded by H_2O.

Clark and Chalmers claim that their variety of EMT can be contrasted with content externalism because the external features cited in their version of externalism:

> [P]lay an ineliminable role – if we retain internal structure but change the external features, behaviour may change completely. The external features here are just as causally relevant as typical internal features of the brain. (Ibid.)

In this way Clark and Chalmers' EMT is argued to be an active externalism—rather than a passive externalism of the content externalist variety—because of the essential role that the external features play in the generation of action. In the next section I begin to consider Clark and Chalmers' arguments for this new form of externalism with an exploration of the distinction between epistemic and pragmatic actions.

Epistemic vs. Pragmatic Actions

A key component of Clark and Chalmers' argument for the extension of cognitive processes and mental states is the distinction between epistemic and pragmatic actions; a distinction originally introduced by David Kirsch and Paul Maglio (1994). Kirsh and Maglio define pragmatic actions as actions that are performed in order to bring one physically closer to a goal. Such actions bring about some change in the environment that is desirable for its own sake. Clark and Chalmers cite the case of putting cement in a hole in a dam as an example because a

situation where the dam is intact is preferable to one where it is leaking and filling the hole with cement brings one physically closer to the goal of having an intact dam. Likewise, my removing the top from a bottle of water when I am thirsty is an example of a pragmatic action because the state of affairs in which I can drink the water from the bottle is one that is desirable to me.

Epistemic actions are defined by Kirsch and Maglio as being: "physical actions that make mental computation easier, faster, or more reliable" (Ibid: 514). Consequently, epistemic actions are not seen merely as the end products of cognition. They are actions that alter the world because some physical change in the environment, which is brought about by the action, "aids and augments" (Clark & Chalmers 1998: 8) cognition in some way. Consider as an example of an epistemic action the process of using a pen and paper to work out a long multiplication sum, such as finding the product of 786 and 345. In a simple case of mathematical reasoning, where the completion of a mathematical operation such as '6 × 6' is required, most of us can just 'see' the answer (Rumelhart et al. 1986 suggest that this is evidence of some kind of pattern-completion mechanism that we possess). However, for the majority of us the solution to questions of longer multiplication does not come so easily. For example, very few of us could just 'see' the answer to 786 × 345. "Instead, we avail ourselves of an external formalism that reduces the bigger task to an iterated series of familiar steps" (Rowlands 1995a: 19). That is, we write down the numbers on a piece of paper and go through a series of pattern completion operations: '5 × 6', '5 × 8' etc., storing the intermediate results according to the long multiplication algorithm. In such a case the task is transformed from being one of finding the product of 786 and 345 to being the task of finding the product of 5 and 6, the product of 5 and 8 and so on. Without this strategy our ability to find the product of the two numbers is severely limited, thus the strategy is one

that aids and augments our long multiplication abilities. So, if I am presented with what in this case ostensibly seems like a cognitive task and I choose to complete this task not by means of mental arithmetic alone but with the help of a pen and paper then, according to Kirsch and Maglio (and also Clark and Chalmers), my doing so is an epistemic action because it aids and augments my ability to complete the cognitive task in question.

Kirsch and Maglio's objective is to unseat the predominant view that all actions are pragmatic actions. "On this view, cognition is logically prior: cognition is necessary for intelligent action, but action is never necessary for intelligent cognition" (Kirsch & Maglio 1994: 526). They hold such a view to be wrong, arguing that not all actions have the sole purpose of bringing one physically closer to a goal. It is their contention that some actions are performed in order to make a problem-solving or cognitive task easier, even though they might demonstrably bring one physically further away from the stated goal. The specifics of their argument consist in presenting evidence of these epistemic actions from studies of subjects playing Tetris to show that certain cognitive and perceptual problems are solved more quickly, more easily, and with less error by acting in the world than by relying solely on computations performed within the head.

It is important to note that although Clark and Chalmers make much of Kirsch and Maglio's distinction between epistemic and pragmatic actions, they do move beyond the arguments proposed by Kirsch and Maglio in one very important respect. Kirsch and Maglio state that epistemic actions:

> [A]re best understood as actions that use the world to improve cognition. These actions are not used to implement a plan, or to

> implement a reaction; they are used to change the world in order to simplify the problem-solving task. (Ibid: 513)

They go on to more precisely clarify epistemic actions as being actions with the function of improving cognition by:

1. reducing the memory involved in mental computation, i.e., space complexity;
2. reducing the number of steps involved in mental computation, i.e., time complexity;
3. reducing the probability of error of mental computation, i.e., unreliability. (Ibid.: 514)

This is the limit of Kirsch and Maglio's claims for epistemic actions: epistemic actions aid and improve performance on certain cognitive tasks. They do not make the stronger claim that epistemic actions are partially constitutive of cognition. This is where the critical difference between their proposal and EMT lies. Clark and Chalmers declare that epistemic actions demand a spread of epistemic credit. Returning to the long multiplication example considered above Clark and Chalmers hold that the use of the pen and paper to complete the cognitive task can be seen as more than a mere add-on or aid to cognition; the use of the pen and paper is part of the cognitive process itself because a portion of the cognitive load is outsourced into the environment. If the use of the pen and paper is essential to the completion of the cognitive task then, they argue, there is no reason to deny that it forms part of the cognitive process. Thus, for Clark and Chalmers, epistemic actions such as these demand a spread of *epistemic credit* across the internal pattern completion processes and the action performed involving pen and paper. This is another way of expressing the core claim of EMT that we saw in the first section of this chapter: that in certain cases external features play

an ineliminable role in the completion of cognitive tasks. So, if we consider a counterfactual situation where internal structure is retained in the form of the pattern completion mechanisms constitutive of my limited multiplication abilities but the external features—in this case my recruitment of pen and paper—are changed then my capacity to successfully complete the cognitive task in question is compromised.

To further demonstrate their claim that epistemic actions demand a spread of epistemic credit, Clark and Chalmers cite the example of the computer game Tetris, first explored by Kirsch and Maglio in their 1994 paper, and it is to Clark and Chalmers' thought-experiment developed from this research that I turn now.

But before continuing we should be aware here of the distinction between the claim that an epistemic actions can form a constituent part of a cognitive processes, and the claim that external features are causally relevant. Clark, in particular, wavers between the stronger constitutive claim and the weaker causal relevance claim. Only the stronger constitutive claim is sufficient to secure the extension of cognitive processes.

Tetris

For the uninitiated, Tetris involves the manipulation of two-dimensional shapes, or 'zoids', on a computer screen which must be rotated so as to fit together to form lines across the screen. The rotation of the blocks on screen is performed at the push of a button. Points are scored when a full horizontal line is completed, this line then disappears. Failure to complete a horizontal line quickly enough will result in an accumulation of falling zoids that cause the game to end once they reach the top of the screen.

Clark and Chalmers ask us to consider three apparently different ways of playing Tetris:

1. The player, sitting before a computer screen, must picture and then rotate the zoids mentally (i.e. without use of the game's controls) in order to determine their fit and align them with the available 'sockets'.
2. The player, instead of mentally rotating the zoids, adopts the standard strategy of physically rotating the image on the screen by pressing a button. This strategy has the benefit of taking less time than the strategy outlined in case 1.[2]
3. The player has a neural implant that performs the necessary rotation of the 'zoids' at the same speed as the computer in the second case.

In the third case the agent has a choice regarding which internal resource to invoke. He can choose to make use of the implant or use 'good old fashioned mental rotation'. The choice exists because each resource demands different brain processes and also places different demands on attention, but as a matter of fact he chooses to utilise the neural implant. Now although each case is clearly different Clark and Chalmers want to argue that each case is also, importantly, on a par with regard to cognitive status. Our intuitions may tell us that in case 1 the rotation of the zoids is performed mentally, whereas case 2 looks, at first pass, like a simple case of non-mental rotation. But case 3 may be harder to classify; our intuitions do not seem to be as clear cut in this case. The computational operations involved in case 3 are the same as those in case 2, yet the computation takes place within the head, just like the (neural) computational operation in case 1. To help guide our intuitions Clark (forthcoming) introduces a fourth case:

[2] Kirsch and Maglio (1994) calculate that the physical rotation of a shape through 90 degrees takes about 100 milliseconds, plus about 200 milliseconds to select the button. To achieve the same result by mental rotation takes about 1000 milliseconds.

4. A Martian whose natural cognitive equipment includes the kind of technology that can fast-rotate the shapes in the same way that the player in case 3, with the aid of a neural implant, can.

In this fourth example our intuitions are clear and we have no hesitation in classifying the case as one of genuine mental rotation according to Clark.[3] He claims, therefore, that case 4 is on a par with case 1. But the computational operations that are involved in case 4 are the same as those involved in case 3. So, says Clark, case 3 is also on a par with case 1 and should be considered as an example of cognitive processing that happens to involve non-neural hardware. To refute this claim the provision of a principled difference between the cases is required. Clark states that if no principled difference can be supplied then case 3, involving the neural implant, is on a par with case 4 and is also, then, on a par with case 1. Furthermore, he continues, since case 2 displays the same kind of computational structure as case 3 it seems that case 2 is also on a par with case 1. The only difference between cases 2 and 3 is that the computational structure is distributed across agent and computer in case 2 instead of internalised within the agent, as in case 3.

Clark and Chalmers argue that if case 3 is on a par with case 1 and the rotation of the shapes, which is carried out by means of the neural implant, counts as a cognitive process, then there does not seem to be any principled reason for denying that the rotation of the shapes in case 2 counts as a cognitive process, or as part of a cognitive process. As already stated; the computational structure is the same in case 2 and case 3 and although the computational structure is internalized in case 3, nothing else of relevance seems different according to Clark and Chalmers. If, they claim, the manipulation of these shapes was carried out

[3] We should beware of the dangers of appealing to intuition. As will become clear in the fourth chapter it is not intuition that will settle the debate in cases like this but the provision of an independently motivated mark of the cognitive or mark of the mental.

in the head we would have no problem in describing the situation as cognitive processing; to deny that doing so physically on a computer screen is a cognitive process, simply because it occurs outside the head, is to beg the question. One cannot point to the skin and skull boundary as a means of justification for the exclusion of this process as part of the cognitive process because that is precisely the question at issue.

Clark and Chalmers go on to make the point that it is not just the presence, in these cases, of advanced external computing resources that raises the issue of the possibility of the extension of cognition into the world. Rather, it is, they claim, a general tendency of human cognizers to: "lean heavily on environmental supports." (Clark & Chalmers 1998: 8). The example considered earlier of the use of pen and paper to complete a long-multiplication task is a good example of just such a tendency. There are, suggest Clark and Chalmers, many other examples of this tendency in action and what is common in each case is that: "[T]he individual brain performs some operations, while others are delegated to manipulations of external media" (Ibid.).

As highlighted in the previous section Clark and Chalmers (and Clark in subsequent writings) are making a stronger claim than that made by Kirsh and Maglio. In their consideration of how experts play Tetris Kirsch and Maglio arrive at the following conclusion:

> We have found that some of the translations and rotations made by players of this video-game are best understood as actions that use the world to improve cognition. These actions are not used to implement a plan, or to implement a reaction; they are used to change the world in order to simplify the problem solving task. (1994: 513).

Clark and Chalmers' conclusions as based on their Tetris thought experiment are different from those arrived at by Kirsch and Maglio. Kirsch and Maglio consider the epistemic actions revealed in their studies as improving the cognitive capacities of a subject by simplifying the problem solving task. And, although they make reference to there being "a tighter coupling of action and cognition" (Ibid: 518) than has traditionally been thought, they nowhere make reference to these actions constituting part of the cognitive process. They maintain that the traditional failure to recognise this potential for an intimate coupling between action and cognition is a mistake and that we need to:

> [R]ecognize that often the point of an action is to put one in a better position to compute more effectively: to more quickly identify the current situation; to more quickly retrieve relevant information; to more effectively compute one's goal. (Ibid: 526)

In this manner Kirsch and Maglio consistently refer to epistemic actions performed in the playing of Tetris as merely *simplifying* the task facing the player or as *improving* the player's cognitive capacities, thus retaining the cognitive processing within the head.

> We have found that in a video-game as fast paced and reactive as Tetris, the actions of players are often best understood as serving an epistemic function: the best way to interpret the actions is not as moves intended to improve board position, but rather as moves that *simplify the player's problem-solving task*. (Ibid: 514, emphasis mine)

> [S]uch epistemic procedures…are not pragmatic procedures; they are procedures that direct the agent to exploit its environment to *make the most of its limited cognitive resources*. (Ibid: 518, emphasis mine).

It seems that there are two different ways of interpreting the results of the Tetris experiments and Clark and Chalmers are keen to push for the stronger constitutive claim by looking for a spread of epistemic credit across internal processes and external actions. They argue for this, as we have seen, by making claims of parity between the different possible strategies adoptable by a player of Tetris. This claim for parity between purely internal processes and hybrid processes that straddle the internal and the external is dubbed the 'Parity Principle' and this forms the focus of the next section.

The Parity Principle

The term 'Parity Principle' is introduced by Clark in *Supersizing the Mind* (2008b), but the original formulation of the principle is given in 'The Extended Mind'.

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head! (Clark & Chalmers 1998: 8)

According to the Parity Principle if a process that is conducted in the external environment contributes to the completion of a cognitive task in a way that were it to go on in the head we would deem it part of the cognitive process, then the process conducted in the external environment should also be seen as part of the cognitive process. Thus, the Parity Principle states that the location of a process is incidental, or even irrelevant, to its potential status as a cognitive process. Processes conducted in the environment should be given equal consideration as potential cognitive processes, or parts thereof, as processes conducted

within the skin and skull. It is not the location of the process which determines its status as cognitive or non-cognitive. It sounds, then, as if the Parity Principle, upon which Clark and Chalmers' active externalism depends, evokes a version of functionalism. This is something that is recognised by Chalmers in his foreword to Clark's *Supersizing the Mind* (2008b). Rejecting the idea that EMT requires functionalism about all mental states Chalmers, nonetheless, concedes that EMT does require a weaker, more attenuated version of functionalism:

> All one needs is the very weak functionalism captured in the Parity Principle: roughly, if a state plays the same causal role in the cognitive network as a mental state, then there is a presumption of mentality, one that can only be defeated by displaying a relevant difference between the two (and not merely the brute difference between inner and outer). (Clark 2008b: xv)

Moreover, if a process that is conducted in the external environment makes an essential—or as they put it, ineliminable—contribution to the completion of a cognitive task then that process, say Clark and Chalmers, should be seen as a constitutive part of the cognitive process. And unless some relevant difference is provided then nothing other than a pre-existing internalist bias would be a reason for rejecting this claim.

Returning once more to look at the Tetris thought experiment we see that Clark and Chalmers' claim is that if the processes described in case 2—the standard Tetris strategy explored by Kirsch and Maglio—were to go on in the head then we would consider these processes as cognitive. But case 3, involving the neural implant, just is, they claim, an instance of the processes described in case 2 going on in the head.[4] If case 3 can be considered an instance of genuine cognitive processing, then, Clark and

[4] See Coleman, forthcoming, for some good reasons why we should reject this claim.

Chalmers argue case 2 must also be considered to be a cognitive process on grounds of parity. Which is to say that if we can accept that the cognitive is not necessarily limited by the neural then it should not matter whether the non-neural element is located within the head or not. If this is correct then the process described in case 2, combining internal processing and action in the world, is a hybrid cognitive process that straddles the internal and the external, thereby extending beyond the boundaries of skin and skull and into the world.

A further example offered in support of these claims goes as follows:

> One can explain my choice of words in Scrabble, for example, as the outcome of an extended cognitive process involving the rearrangement of tiles on my tray. Of course, one could always try to explain my action in terms of internal processes and a long series of "inputs" and "actions", but this explanation would be needlessly complex. If an isomorphic process were going on in the head, we would feel no urge to characterize it in this cumbersome way. In a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of *thought*. (Clark & Chalmers 1998: 9-10).

Given their claims here regarding the isomorphism between internal and extended processes, one could, perhaps, be forgiven for thinking that the Parity Principle makes a prima facie claim for similarity (rather than parity) between internal processes and certain combinations of internal and external processes. But Rowlands (2009a) warns against reading the parity principle in this way, since to do so expresses a crude and mistaken view of EMT. It is futile to object, on this view, that internal processes are demonstrably different from the external ones simply in virtue of the fact that internal processes involve complicated neural happenings whereas the environmental processes do not since the parity expressed in

the Parity Principle is pitched at a level of abstraction above that of brute similarity in processing. The claim was never that the coupling of internal and external processes constitutes an externalisation of what is already in the head. Rather, the claim is that *were such a process to go on in the head* then it would be cognitive.

Nonetheless, given the possibility that a mistaken similarity-based interpretation of the Parity Principle affords for making the sort of objection sketched above, some proponents of EMT have argued that relying on the Parity Principle is a mistake. Richard Menary (2006), for example suggests that the Parity Principle, as currently framed, is flawed and should be dropped by advocates of EMT. It is flawed because it emphasises equivalence between internal and external processes, whereas Menary argues that it is pertinent to emphasise the differences.

Menary prefers the term 'Cognitive Integrationist' over 'Extended Mind Theorist' since he sees the task as being one of integrating disparate internal and external processes into an extended cognitive system. Menary claims that the integration of internal and external elements of cognitive processing makes more sense when we drop the Parity Principle and understand arguments in favour of EMT as integrating apparently disparate processes into a cognitive whole for the purpose of completing certain cognitive tasks. In this case the external processes that constitute part of the extended cognitive process may be quite different from the internal processes with which they are coupled. Indeed it is in these differences that the capacity to aid and augment cognition lies according to Menary. He claims that it is important to the integration of internal and external processing that external manipulations do something different from brain processes. These differences are important because they can help to transform our abilities to complete cognitive tasks and

can, consequently, potentially improve our abilities to complete them.[5] Thus, for Tetris gamers the very fact that the processes required to rotate the zoids on-screen differ from those required to do so mentally means that the task can be achieved more quickly and with a greater degree of success.

It is precisely because of the difference between internal and external processes that an organism can complete certain novel cognitive tasks (see Menary 2006 & Sutton 2006). Integration of the external manipulation of information-bearing structures with internal processing expands an organism's cognitive capacity. Thus, the differences between internal and external processing are more important than any similarities we may invoke, according to the integrationist perspective. We can view cognitive processes as being hybrid; that is, composed of both internal and external processing components because of the fact that the external/extended processes can do things that the internal processes cannot, or do not, do. Therefore, Menary emphasises that external structures and processes possess very different properties from internal ones:

> In the head, there are connectionist vehicles and processes over them. There isn't anything in the environment that looks like connectionist vehicles and processes over them. There are symbols, such as diagrams and linguistically structured vehicles and their manipulation is different from the manipulation of connectionist vehicles. (Menary 2006: 342).

And Mark Rowlands agrees:

> External structures and processes possess quite different properties from internal ones; and it is precisely this difference that affords the cognitive

[5] This is also expressed by Rowlands in his 'Transformation Thesis' 1998: 330.

> agent the opportunity to accomplish certain tasks that it could not, or might not, be able to accomplish purely by way of internal cognitive processes. Without these differences, the external processes would be *otiose*. (Rowlands 2009a: 3-4).

The enhanced possibilities for manipulation and exploitation that are afforded cognitive agents because, for example, of the stability of external structures, are argued not to be available for internal processing alone. Thus, the enhanced possibilities provided by the resources of the external environment and the potential for acting on them underwrite the ability of the cognitive agent to accomplish novel cognitive tasks; cognitive tasks that the agent would not otherwise have been able to complete given the use of internal processes alone.

According to the cognitive integrationist case 1 and case 2 of Clark and Chalmers' Tetris thought experiment can both be considered as cognitive processes. This is not, however, because case 1 is *similar* to case 2 but because the player in case 2—in conjunction with the manipulation of the external information-bearing structures of the game's hardware—constitutes an integrated cognitive system, a possibility which is afforded the player precisely because the external information-bearing structures differ from the internal provisions.

So the integrationist perspective rejects the use of the Parity Principle as a means for arguing for the extension of cognition into the environment and seeks to focus, instead, on the differences between internal and external resources. But such a move cannot be quite right. The integrationists are correct to emphasise the differences between internal and extended putative cognitive processes and in this regard, the integrationist perspective provides a means of countering many of the objections levelled at EMT that take the Parity Principle to be introducing

a brute similarity-based criterion for judging whether or not an external process counts as part of a genuine cognitive process. But there must, still, be *something* that both sets of processes have in common, otherwise what is there to mark out the cognitive from the non-cognitive?[6]

The mistaken similarity-based reading of the Parity Principle states that if an external process is enough like an internal cognitive process then there is no reason to deny the external process cognitive status as part of a larger, extended cognitive process. But such a reading of the Parity Principle looks to have been misguided from the start. A similarity-based criterion was not what Clark had in mind when the Parity Principle was introduced:

> [F]ar from requiring any deep similarity between inner and outer processes, the parity claim was specifically meant to undermine any tendency to think that the shape of the (present day, human) inner processes sets some bar … on what ought to count as part of a genuinely cognitive process. The parity probe was thus meant to act as a kind of veil of metabolic ignorance, inviting us to ask what our attitude would be if currently external means of storage and transformation were, contrary to the presumed facts, found in biology'. (Clark, forthcoming)

Clark stresses, then, that the Parity Principle does not make any claims about identity at the level of processing at all. And this is reinforced by what was said originally in 'The Extended Mind', particularly with regard to language.

> Without language, we might be much more akin to discrete Cartesian 'inner' minds, in which high-level cognition relies largely on internal resources. But the advent of language has allowed us to spread this

[6] This issue will surface again in discussion of the Mark of the Cognitive in chapter 4.

> burden into the world. Language, thus construed, is not a mirror of our inner states but a complement to them. It serves as a tool whose role is to extend cognition in ways that on-board devices cannot. Indeed, it may be that the intellectual explosion in recent evolutionary time is due as much to this linguistically-enabled extension of cognition as to any independent development in our inner cognitive resources. (Clark & Chalmers 1998: 18)

Interpretations which take it that the Parity Principle holds that external processes must be similar to internal processes in order to count as cognitive are mistaken. The Parity Principle is not the mere postulation of similarity between internal and extended processes. However, clearly there must be identity at *some* level between case 1 and case 2 in the Tetris thought experiment as both are said to be cases of genuine cognitive processing. Thus, Clark and Chalmers pitch the Parity Principle at a level of abstraction higher than that of what implements the process. Rather, the parity that can exist between internal and extended processes is parity of the sort mentioned in David Chalmers' quote above, which is to say that two processes are analogous according to the Parity Principle if they fulfil the same causal role within a cognitive system. This does not entail that the two cases must be the same, there can and perhaps must be genuine differences between the two cases but these differences do not matter.

Or, rather, these genuine differences do matter, as the integrationist perspective correctly points out, but not as a basis for denying cognitive status to extended systems. Menary seems to echo Clark's sentiments here; he suggests that objectors have attacked a flawed version of the Parity Principle:

> [I]t has become something of an albatross around Clark's neck. Internalists have seized upon the parity principle, attacking a *flawed*

> comparative version which, schematically, can be stated as: if an external process/vehicle X plays a similar role to an internal process/vehicle Y (where Y is cognitive), then X is cognitive. (Menary 2006: 333, emphasis mine)

So, it seems that the Parity Principle, despite the initial protestations of the integrationists, can be salvaged but only so long as the differences predicted and required by the integrationist perspective are appreciated. We will examine in later chapters some possible further problems with the Parity Principle but for now I move on, armed with an understanding of the Parity Principle informed by the integrationist perspective to consider how Clark and Chalmers develop their argument based on the Parity Principle with regard to cognitive processes and I do so by examining the case of Otto; who has become the poster boy of EMT.

Otto

So that we can better appreciate Clark and Chalmers' argument in favour of the extension of cognitive processes into the world, they offer us the comparison of Otto and Inga. Otto has Alzheimer's disease and consequently his biological memory is not as reliable as it used to be. Otto carries a notebook with him wherever he goes and when he learns some new information he writes it in the notebook. Whenever he needs to access some old information, he looks it up in the notebook. Now, suppose that Otto hears about an exhibition at the Museum of Modern Art and he has a desire to see it; he looks in his notebook, which tells him that the Museum is located on 53rd Street, and off he goes. The question here is whether or not Otto's use of the notebook counts as an example of a genuine cognitive process, specifically, one of remembering.[7]

[7] In the original treatment Clark and Chalmers ask us to consider the case of Otto for the purpose of presenting an argument in favour of the extension of mental states into the world, specifically beliefs. In the meantime, the example of Otto has been appropriated

Inga, unlike Otto, does not have Alzheimer's disease and her biological memory is in fine working order. Nevertheless, say Clark and Chalmers, if we compare the cases of Otto and Inga, we will find that the cases are analogous, as per the Parity Principle. For Otto the notebook plays the same role that ordinary biological memory plays for Inga: Otto has come to rely on his notebook in the same way that Inga relies on her biological memory. For Otto, the notebook is a constant just as Inga's memory is a constant for her. Although there are various immediately apparent differences between the two cases Clark and Chalmers argue that these are merely shallow differences and to focus on them is to miss the bigger picture. In order to deny cognitive status to Otto's use of the notebook an opponent needs to show that Otto's case is different from Inga's in some important and relevant respect.

Once more, as emphasised by Chalmers (see Clark 2008b: xv), the relevant difference cannot simply be that for Inga the information is stored in the head whereas for Otto the information is stored in the notebook. But perhaps it could be argued that Inga has more reliable access to her memories than Otto has? Someone could potentially steal Otto's memory store (notebook) after all; something which would prove considerably more difficult in Inga's case. Clark and Chalmers agree that reliability and constancy *are* relevant here. In order for Otto's notebook to count as part of the cognitive process of remembering it needs to be constantly reliably available to him. And in this case it is. Otto always carries the notebook with him and when asked a question he will not answer that he does not know until after he has consulted it. Otto's case would not count as a case of genuine remembering if he were merely

for arguments both in favour of and against the extension of cognitive processing. A discussion of the possibility of the extension of mental states follows later.

accessing information from a guide-book or encyclopaedia as a one-off, for example.

Thus, Clark and Chalmers argue, Otto's notebook is coupled with him so as to realise an extended cognitive system precisely because it is a constant for him. The information is always reliably available for him when needed. Sure, the notebook *might* get stolen or lost, and Inga's memory seems, by contrast, to be a lot more secure. But the difference in this case is one of degree and not one of kind (Rowlands 2003b). A surgeon could remove the relevant part of Inga's brain associated with her long-term memory, she could fall and damage her brain in hitting her head, or a skilled hypnotist could plant apparently veridical, yet false, memories in her. It is true that Otto's external memory store is not perfectly reliable, but neither is Inga's internal memory store. So an objection on the grounds of reliability is not sufficient to discount Otto's case. What matters, say Clark and Chalmers, is that: "the information is easily available when the subject needs it, and this constraint is satisfied equally in the two cases" (Clark & Chalmers 1998: 15). Thus, one criterion that must be met by an external, or non-biological, putative cognitive resource according to Clark is:

1. That the resource be reliably available and typically invoked. (Clark 2008b: 79)

An opponent of EMT might press the point and question the difference in access that each has to their respective 'memory stores'. Otto's access to his notebook is not constant in the same way that Inga's is. That is, it comes and goes depending on the situation. He showers without his notebook, for example, and cannot read the information contained in it when it is dark. Inga, by contrast, it might be claimed, always has her memory store with her and thus always has access to it. But this is not

strictly true, Inga's access to her memory store also comes and goes; when she is asleep or when she is intoxicated, for example. Again, the difference here is one of degree only and is not sufficient to form an objection to discount Otto's case as one of genuine belief.

> Temporary disconnections do not seem to be decisive, what is crucial is that the information is easily available when the subject needs it, and this condition is met equally in both cases. (Rowlands 2003b: 181)

Perhaps the relevant difference, instead, is that Inga has *better* access to her memories than Otto does?

> Inga's "central" processes and her memory probably have a relatively high-bandwidth link between them, compared to the low-grade connection between Otto and his notebook (Clark & Chalmers 1998: 15).

Again, Clark and Chalmers argue that this is not a relevant difference. They claim that we can construct an analogous case whereby Lucy, a friend of Inga, has only a low grade connection between her biological memory and central systems. This could be due to Lucy's non-standard biology, or it could be due to her past misadventures. Either way, what counts, again, is that the information that Lucy requires is accessible; even if the processing is less than efficient. This applies equally to Otto's case. A second constraint on a putative non-biological cognitive resource offered by Clark then is:

> 2. That information contained in the resource should be easily accessible as and when required. (Clark 2008b: 79).

A further objection considered by Clark and Chalmers states that Otto's and Inga's respective *access* to the information stored in memory differs

in an important detail. Inga has direct access to her memories, via introspection, say; whereas Otto's access to his memories comes only by way of perception and perceptual access to information in the environment has a distinct phenomenology, but it is not clear, say Clark and Chalmers, why this should make a difference to the status of Otto's belief. What Clark and Chalmers are proposing is that Otto and his notebook be considered as an extended cognitive system and that within this system the flow of information between Otto's notebook and his brain is not perceptual. It is not perceptual, they claim, because it does not involve the impact of something that is outside the system. In this way the flow of information between the notebook and the brain is more like the flow of information within the brain in normal cases. In Otto's case there is a distinct perceptual phenomenology that accompanies the retrieval of the information from the notebook, which is not there in Inga's case, but it is not clear, argue Clark and Chalmers, that this counts as a relevant difference.

We can, perhaps, imagine a case where a cognitive system has access to its memory that has a distinct perceptual phenomenology. Suppose, for example, that there exist aliens whose biological routines store bit-mapped images of printed words. Access to this information, let us suppose, is via bit-mapped signals sent to the visual cortex. Would this bit-mapped storage form part of the alien's cognitive system? Would we discount the 'memories' of Arnold Schwarzenegger's Terminator as genuinely cognitive because they are displayed in his visual field and thus have a distinct perceptual phenomenology associated with them?

Tied in with this worry is the concern that normal biological memory, like Inga's, is automatically endorsed. A third constraint proposed by Clark on a putative non-biological resource forming part of a genuinely cognitive process, then, is:

3. That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory (Ibid: 79).

Taking these three criteria offered by Clark we can see that a book in the library would not generally count as an extension to cognition, nor would ordinary mobile access to an internet search-engine, such as Google. Otto's notebook, however, does count as an extension to his cognition. Thus, conclude Clark and Chalmers, Otto's case gives us an example of a genuine extended cognitive system. The argument proceeds from the Parity Principle considered earlier.

> Certainly, insofar as beliefs and desires are characterized by their explanatory roles, Otto's and Inga's cases seem to be on a par: the essential causal dynamics of the two cases mirror each other precisely (Clark & Chalmers 1998: 13).

The claim is that Otto's case is the same as Inga's case in all important respects and the challenge for the opponent of EMT is to show that this parity is defeated by citing some relevant and important difference between the cases. Furthermore, the Otto case incorporates the lessons regarding the Parity Principle that we learned from the cognitive integrationists in the previous section. Given that Otto has Alzheimer's disease his biological memory lacks the stability or reliability that it used to have. A notebook can provide this, thus the *difference* between the internal structures and processes and the external ones are important here. To put it crudely, the notebook provides Otto with something that his brain cannot. It is for this reason that Otto chooses to place some of the cognitive load here. So the differences between the inner and extended

processes do matter and are important, as the integrationist perspective correctly points out, but not, claim Clark and Chalmers, as a basis for denying cognitive status to extended cognitive systems.

Clark and Chalmers contend, then, they have demonstrated that cognitive processes can extend beyond the boundaries of the individual by being partially constituted by features of the environment. But it is not clear that Clark and Chalmers have provided a conclusive case for constitutive parts of cognitive processes extending into the world. Their argument proceeds by stating that if we accept the Parity Principle then we have no good reason to reject Otto's notebook as an external memory store.

But an internalist may accept the Parity Principle and reject the argument that Otto's notebook should count as an extended memory. Just as the external location of a process should not automatically deny it cognitive status, the internal location of a process should not automatically award it cognitive status. Thus, the internalist might argue that if the extended process involving Otto's notebook were to go on in the head, we would still deny it cognitive status. What is required here to settle the issue is a definition of what a cognitive process is; a mark of the cognitive.[8] Without it the internalist can argue that an equally valid interpretation of the case is that Otto's notebook transforms the nature of the cognitive task facing Otto while the genuinely cognitive processing is internal. The EM theorists, says the internalist, have given us no good reason to see the external processes as anything other than aids to cognition and the genuinely cognitive as remaining skull-bound.

However, we can put these worries to one side for the moment since Clark and Chalmers seek to move beyond these arguments for the

[8] This is key issue for EM theorists. I demonstrate why in more detail in chapter 4.

extension of cognitive process to argue that mental states can also be partially constituted by features of the environment. This move is key, since as will become clear, EMT proper requires the extension of mental states; the extension of cognitive processes alone is not sufficient to secure EMT. With that in mind it is to these arguments for the extension of mental states that I now turn and this requires further analysis of the cases of Otto and Inga.

Mental States

Having offered their arguments in favour of the extension of cognitive processes into the environment, Clark and Chalmers take their externalism a stage further. They speculate that someone might object that more than mere processing is required for mentality and so, regardless of whether or not cognitive processes are partially constituted by environmental factors, the *mind* remains firmly rooted within the boundaries of skin and skull.[9]

> So far we have spoken largely about "cognitive processing", and argued for its extension into the environment. Some might think that the conclusion has been bought too cheaply. Perhaps some *processing* takes place in the environment, but what of *mind*? Everything we have said so far is compatible with the view that truly mental states - experiences, beliefs, desires, emotions, and so on - are all determined by states of the brain. Perhaps what is truly mental is internal, after all? (Clark & Chalmers 1998: 12)

Clark and Chalmers want to challenge this claim that mental states are necessarily internal. They propose that mental states, in addition to

[9] Unfortunately, Clark and Chalmers don't tell us what the something more than processing that constitutes mental states might be. The issue here is the mark of the mental will be discussed in detail in chapter 4.

cognitive processes, can extend into the environment and their argument for this focuses on beliefs: they argue that beliefs can literally extend into the world.

> While some mental states, such as experiences, may be determined internally, there are other cases in which external factors make a significant contribution. In particular, we will argue that *beliefs* can be *constituted* partly by features of the environment, when those features play the right sort of role in driving cognitive processes. *If so, the mind extends into the world*. (Ibid: 12, emphasis mine)

In support of the claim that beliefs extend into the world we are asked to compare Otto's case with that of Inga, as above. Inga hears about an exhibition at the Museum of Modern Art and has a desire to see it. She thinks for a moment before recalling that the Museum of Modern Art is on 53rd Street and heading off. Why does Inga go to 53rd Street? Because she has a *desire* to see a particular exhibition in the Museum of Modern Art and she *believes* that the Museum of Modern Art is on 53rd Street. Inga's case is best described in terms of personal level mental states and contents.

Now, consider the status of Inga's belief. Did Inga acquire a new belief, the belief that the Museum of Modern Art is on 53rd Street, having consulted her memory? Did Inga re-acquire that belief, having held it previously, after consulting her memory? Or, did Inga believe that the Museum of Modern Art was located on 53rd Street before she consulted her memory? It seems that the latter answer is the correct one. Inga, from when she first learned the location of the Museum of Modern Art, held the belief that it was on 53rd Street. This belief became occurrent when she accessed her memory store but she doesn't stop believing that the Museum is on 53rd Street when the belief is no longer an occurrent one. It

is not necessary that a particular piece of information be constantly present before consciousness in order for a subject to believe it.[10]

Next, we compare Inga's case with that of Otto. Suppose Otto hears about the same exhibition in the Museum of Modern Art and he too has a desire to see it. He looks in his notebook, which tells him that the Museum is located on 53rd Street, and off he goes. Why does Otto go to 53rd Street? Clark and Chalmers claim that we can apply the same reasoning to Otto's case as we do to Inga's. Otto has a *desire* to see a particular exhibition at the Museum of Modern Art and *believes* that it is located on 53rd Street. We can also apply the same questions regarding the status of Otto's belief as we did to Inga's belief. Does Otto acquire a new belief, the belief that the Museum of Modern Art is on 53rd Street, having consulted his notebook? Does Otto re-acquire that belief, having held it previously, after consulting his notebook? Clark suspects that all of these options will prove to be the wrong ones, highlighting the functional role that the notebook plays in Otto's everyday life.

> [I]f we follow Otto around for a while, we will see how unnatural this way of speaking is. Otto is constantly using his notebook as a matter of course. It is central to his actions in all sorts of contexts, in the same way that an ordinary memory is central in an ordinary life. (Ibid: 13)

Alternatively, however, we could say that Otto's case is best described by stating that the beliefs that are properly attributed to Otto are ones such as 'Otto believes that the location of the museum of modern art is written in

[10] Since Clark and Chalmer's argument for the extension of mental states focuses on dispositional beliefs a lot hangs on how we understand the status of dispositional beliefs. Some may argue that Inga does *not* continue to believe that the Museum is on 53rd street when the belief is no longer occurrent, and that dispositional beliefs are beliefs that we are disposed to have, i.e. to form, counterfactually, under such and such conditions. Alternatively one might simply reject the idea that dispositional or standing beliefs are part of the mind (e.g. Gertler 2007).

the notebook'. This strategy is one that Clark calls the two-step objection, and it goes like this:

> [A]ll Otto actually believes (in advance) is that the address is in the notebook. That's the belief (step 1) that leads to the looking (step 2) that then leads to the (new) belief about the actual street address. (Clark, forthcoming).

Clark's response to this objection is to question why we don't apply the same reasoning to Inga's case. That is, why don't we say that all Inga actually believes in advance is that the information she requires is stored in memory. This is the belief (step 1) that leads to the retrieval (step 2) that then leads to the new belief about the actual street address. The reason that we don't apply this 2-step strategy in the case of Inga is that it seems to add, in the words of Clark; "spurious complexity" (Ibid.). The best (or at least the standard) way of describing Inga's case is by saying that Inga wants or desires to go to the Museum of Modern Art and believes that the Museum is located on 53rd Street. To employ the two-step strategy to describe Inga's case seems unnecessary because she doesn't rely on beliefs about her memory to guide her everyday actions—the use of her memory is generally automatic and transparent. Clark claims that the same can be said for Otto:

> Otto is so accustomed to using the book that he accesses it *automatically* when bio-memory fails. Calls to the notebook are as deeply and *subpersonally integrated* into his problem-solving routines as calls to external rotation for the expert Tetris players. The notebook has become *transparent equipment* for Otto, just as biological memory is for Inga. And in each case, doesn't it add needless and psychologically unreal complexity to introduce additional beliefs about the book or biological memory into the explanatory equations? (Clark 2008b: 80, emphasis mine).

So, in Inga's case we don't employ the 2-step strategy because it needlessly complicates the matter. And it does so because Inga's access to her memory is automatic and transparent; her access is subpersonal, it is not something that she need have conscious control over. Similarly then, it is claimed that Otto's access is automatic and transparent. Because he has grown to rely on the notebook his access to it has 'gone subpersonal' and thus the 2-step strategy is to add one step too many to our explanation of his behaviour and associated mental states.

If Otto's access was *not* subpersonalised in this way then presumably it would be correct to invoke the 2-step strategy as his access would not then be automatic and transparent. Consider, for example, that I ask someone for the location of the Museum of Modern Art and write the address down in a notebook. Later, I look the information up in the notebook when I decide to go to an exhibition at the museum; in such a case I would best be described as believing that the location of the Museum of Modern Art is written in my notebook, not that the Museum of Modern Art is located on 53rd street, or that I know the location of the Museum of Modern Art. I must make a conscious decision to look in the notebook in order to retrieve the information that I require—my access to the notebook is not automatic and transparent in the same way that Otto's is. However, if I choose to employ this strategy over the long term then as I begin to use the notebook more and more, the idea is that my access to the notebook may become automatic and transparent and at this stage the 2-step strategy would no longer be the correct strategy to employ in describing my actions and attributing mental states. My access to the notebook becomes fluid and immediate. The notebook and my accessing of the notebook are said to become part of the subpersonal architecture of my personal level contents. This is also the case for Otto, but because of

his condition he *has* to rely on the notebook, consequently, adaptation would presumably be quicker and easier.[11]

This kind of case has parallels with the commissurotomy and callosal agenesis cases considered by Susan Hurley (1998b), which we will examine in more detail in chapter three. A commissurotomy is a procedure whereby the corpus callosum—connective tissue linking the left and right hemispheres of the brain—is severed, often with the purpose of reducing or eliminating seizures. Callosal agenesis on the other hand is a congenital condition where an individual is born without the corpus callosum; these individuals are often referred to as acallosals. Commissurotomy patients are faced with the challenge of integrating information between the two hemispheres of the brain as the normal method of information integration has been removed. These cases are interesting to Hurley because of the relevance they have to questions about the unity of consciousness and, in particular, for inducing, under experimental conditions, what Hurley calls partial unity.

Partial unity may be described as falling midway on a scale between a normally functioning human being with a single conscious perspective and an individual exhibiting dissociative identity disorder, which is to say that a partially unified individual does not exhibit distinct dissociated personalities but that inconsistencies in conscious content in separate brain hemispheres can be induced under experimental conditions. The problem of information integration between hemispheres does not really arise for an acallosal as she will always, of necessity, have employed external methods of information integration—such as using cross-cuing

[11] If subpersonal, automatic and transparent access is a requirement then it is not clear that the Otto plus notebook example is the best choice for promoting EMT. Is it really plausible to suppose that Otto uses his notebook without consciously thinking about it? And, in any case, is there any reason to suppose that one cannot quite consciously try to access normal biological memory?

and access movements—between the two hemispheres and is, consequently not best described as being partially unified. In this case the external means of integration is a pre-condition of the unity of consciousness. Newly operated commissurotomy patients are, on the other hand, initially best described as disunified, or partially unified if they employ external methods of integration. However, as these commissurotomy patients continue to practice this external means of integration, the action becomes automatic and transparent. It goes subpersonal. When this happens we should no longer attribute disunity to the commissurotomy patient. So it would be inappropriate to call commissurotomy patients disunified at the personal level given a number of years of making use of access movements and cross-cuing to realise unity. In these cases, with internal methods of integration no longer possible, external methods of integration are a pre-condition for unity at the personal level.

Clark and Chalmers presumably consider Otto's case similar because internal methods of memory storage are no longer possible for him, at least not with any degree of reliability. Thus the external storage he now employs becomes a pre-condition for the successful formation of new beliefs and if his access to the notebook is automatic and transparent the two-step objection is thought to fail.[12]

So Clark and Chalmers argue that with regard to Otto, even before he looks up the location in his notebook, the attribution of the belief *that the Museum of Modern Art is on 53rd Street* makes the most sense. It is not

[12] Once again, in comparison with a commissurotomy patient who subpersonally integrates information between the two hemispheres of the brain, it may not seem plausible that Otto could use his notebook to retrieve information without it being under some measure of conscious control. As we will see in chapter 5 though, the argument in support of Otto's belief extending seems to rest more on the functional role that the notebook plays for Otto rather the nature of his access to it. It may be the case that the automatic and transparent access requirement is introduced purely to deal with the two-step objection.

necessary for a particular piece of information to be constantly present before consciousness in order for it to be a belief. So, as in Inga's case, Otto's belief becomes an occurrent one when he accesses his memory store, the notebook.

If Otto's case warrants the attribution of the belief *that the Museum of Modern Art is on 53rd Street* even before he looks in his notebook then, Clark and Chalmers claim, Otto's belief extends into the world. Otto's belief extends into the world because the memory, on which this belief depends, extends into the world as it is located in Otto's notebook. Clark and Chalmers claim that analogously to Hilary Putnam's Twin-Earth thought experiment we can construct a Twin-Otto thought experiment which supports this conclusion that beliefs can and do extend beyond the brain. Imagine that on Twin-Earth there is a Twin-Otto who is exactly like Otto in every respect except for the fact that instead of writing in his notebook that the Museum of Modern Art is on 53rd Street, he mistakenly writes that the Museum is on 51st Street.

> Today, Twin Otto is a physical duplicate of Otto from the skin in, but his notebook differs.[13] Consequently, Twin Otto is best characterized as believing that the museum is on 51st Street, where Otto believes it is on 53rd. *In these cases, a belief is simply not in the head*. (Clark & Chalmers 1998: 14)

Here, the difference between Content Externalism and Active Externalism emphasised by Clark and Chalmers becomes manifest. A Twin-Otto with the belief *that water* (xyz) *is wet* will behave no differently from Otto who believes *that water* (H20) *is wet*, whereas

[13] In making this claim Clark and Chalmers may be committing the error of assuming the possibility of general duplication: that internal states can, in general, be duplicated in counterfactually altered environments. In chapter 3 we will examine whether or not the duplication required for their thought experiment to work can be achieved as easily as they hope.

Twin-Otto in the above case will go to 51st street and Otto will go to 53rd street. Otto's notebook is, it seems, an active and proximal external feature that is causally relevant in the production of behaviour. The notebook for Otto (and for Twin-Otto) plays an ineliminable role in the production of behaviour and in the formation and subsequent attribution of mental states—without it we could not attribute certain beliefs and memories to Otto. So, given his use of the notebook, Clark and Chalmers argue that Otto is best described as believing *that the Museum of Modern Art is on 53rd Street*. But there's no doubt that Otto comes to believe this at some point. The real question is whether or not Otto believes *that the Museum of Modern Art is on 53rd Street* before he looks in his notebook. Clark and Chalmers take themselves to have demonstrated that Otto knows where the Museum of Modern Art is before he looks it up—just as Inga does before she brings it to mind—in which case the notebook constitutes part of the belief state and Otto's belief extends.

But, once again it is not clear that Clark and Chalmers have done enough to secure this conclusion. It is possible to accept the parity principle and still reject the claim that Otto knows the location of the Museum before he looks in his notebook. If Otto's looking in his notebook does not count as a genuine case of remembering, regardless of its location, then the notebook will not constitute part of Otto's belief state.

It seems that all that Clark and Chalmers have conclusively demonstrated is that the notebook has a key role to play in the formation of Otto's belief *that the Museum of Modern Art is on 53rd Street*, rather than the stronger claim that it is a constitutive part of his belief state. The difference between these two claims is the difference between EMT and EMMT: the Extended Machinery of Mind Hypothesis. I can agree that Otto believes that MOMA is located on 53rd street, and I can hold that this is true because of his notebook, without adhering to the claim that

Otto's belief extends into the world. The notebook forms part of the mechanism that *enables* Otto to believe that MOMA is located on 53rd street, but it is not a constitutive part of his belief. We will examine exactly why this should be so in chapter 4 but, for now, I just want to highlight the tension that exists, particularly in Clark's writings, between the claims made for EMT and the arguments themselves, which make a good claim for EMMT but do not decisively prove the stronger claim. This is the difference between mechanisms that enable cognition and mental states and mechanisms that constitute cognition and mental states. In the next section I proceed to highlight where this tension is evident in Clark's Active Externalism.

Constitution vs. Enabling

Analysis of Clark and Chalmers' 'The Extended Mind' has revealed that they frequently make claims for the partial external constitution of mental states and cognitive processes – for their extension into the environment. For example:

> Once we recognize the crucial role of the environment in constraining the evolution and development of cognition, we see that *extended cognition* is a core cognitive process, not an add-on extra. (Ibid: 12)

> In particular, we will argue that *beliefs can be constituted partly by features of the environment*, when those features play the right sort of role in driving cognitive processes. If so, *the mind extends into the world*." (Ibid.)

> Thus, in seeing *cognition as extended* one is not merely making a terminological decision; it makes a significant difference to the methodology of scientific investigation. (Ibid: 10)

> Another example may be language, which appears to be a central means by which *cognitive processes are extended into the world*. (Ibid: 11)
>
> In these cases, a belief is *simply not in the head*. (Ibid: 14, all emphases mine)

Such a view is also evident in the Parity Principle, the claim on which their argument for EMT is based:

> If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then *that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!* (Ibid: 8, emphasis mine)

But we also questioned whether these conclusions about the partial external constitution of mental states and cognitive processes are warranted. In his subsequent writings Clark has tended, for the most part, to steer clear of the sorts of expressions whereby he talks of the extension of cognition, cognitive processes, or mental states into the world. Instead, we get references to the extension of the material vehicles of mind, or to the machinery of mind or cognition. Dubbing his version of EMT as EXTENDED in his latest book, Clark characterises it as follows: "The physical mechanisms of mind, EXTENDED suggests, are simply not all in the head" (Clark 2008b: 85).

But, frequently running alongside these expressions about the extension of the machinery of mind are bolder claims about the extension of cognition and mind itself:

> The local mechanisms of mind, if this is correct, are not all in the head. *Cognition leaks out into the body and world.* (Clark 2008b: xxviii, emphasis mine)

> Such body- and world-involving cycles are best understood, or so I shall argue, as quite literally extending the machinery of mind out into the world—as building extended cognitive circuits that are themselves the minimal material bases for important aspects of human thought and reason. *Such cycles supersize the mind.* (Clark, 2008b xxvi, emphasis mine).

And, in discussing the famous exchange between the physicist Richard Feynman and the Historian Charles Weiner about whether Feynman's notes constituted his work or merely a record of his work Clark remarks:

> …I would like to…suggest that Feynman was actually *thinking* on the paper. The loop through pen and paper is part of the physical machinery responsible for the shape and flow of his thoughts and ideas that we take, nonetheless, to be distinctively those of Richard Feynman. It reliably and robustly provides a functionality which, were it provided by goings-on in the head alone, we would have no hesitation in designating as part of the cognitive circuitry. (Ibid: xxv)

Within these lines we can distinguish two views. The first of these is EMT; the view that Feynman was thinking on the paper, the view that Otto's belief extends into the world in the form of his notebook, the view that pushing buttons to rotate blocks on a computer screen is partially constitutive of a cognitive processes. The second view is EMMT; the view that the machinery of mind extends into the world. This is a sort of enabling-externalism which views Otto's notebook as a belief-forming mechanism without which Otto would not be able to form the belief *that*

the Museum of Modern Art is on 53rd street but which doesn't imply that he has the belief before consulting his notebook.

As discussed earlier, such a view is the one that Kirsch and Maglio seem to take with regard to their examination of epistemic actions in relation to playing Tetris, which is to say that without the use of epistemic actions players would find it a lot harder to complete the rotation and alignment of the zoids with such speed. We also examined how Clark and Chalmers seek to go further than Kirsch and Maglio in arguing that epistemic actions demand a spread of epistemic credit, claiming that the action in question is partially constitutive of a cognitive process and not merely an enabler or facilitator of a cognitive process. Thus, it is EMT rather than EMMT that Clark and Chalmers try to defend in 'The Extended Mind'. However, alongside their claims for EMT in this paper are expressions of a position that is more akin to EMMT:

> Moreover, it may be that the biological brain has in fact evolved and matured in ways which factor in the reliable presence of a manipulable external environment. It certainly seems that evolution has favoured on-board capacities which are especially geared to parasitizing the local environment so as to reduce memory load, and even to *transform the nature of the computational problems themselves*. (Clark & Chalmers 1998: 11, emphasis mine)

> It is not just the presence of advanced external computing resources which raises the issue, but rather the general tendency of human reasoners to *lean heavily on environmental supports*. (Ibid: 8, emphasis mine)

Thus there seems to be a tension in the original paper. The tension is between EMT as advertised—which is as nothing less than an argument for the extension of mind into world—and the reality, which is that

arguments for EMT secure only EMMT, a thesis about the physical mechanisms that enable cognition and the formation of mental states.

> The answer, we claimed, was that mental states, including states of believing, could be *grounded* in physical traces that remained firmly outside the head. (Clark, forthcoming, emphasis mine)

That will suffice to highlight that this tension, between Active Externalism as a version of EMT and as a version of EMMT exists. In chapters four and five a more detailed examination of the distinction between EMT and EMMT will be provided along with a demonstration of how EMT as originally formulated by Clark and Chalmers, collapses into EMMT. Next, however, I continue the expository work by examining the work of Mark Rowlands and Susan Hurley, who are taken to be two of the most foremost exponents of EMT.

Chapter 2

Introduction

In this chapter I continue the expository work on EMT started in chapter one, with the focus switching to Mark Rowlands. Rowlands' version of EMT was initially labelled Environmentalism and he later comes, in *Externalism* (2003b), to endorse a variety of Vehicle Externalism. I examine both theses in order to show that —as with Clark and Chalmers' Active Externalism— Rowlands' work fails to provide conclusive support for an interesting variant of EMT. Crucially, Rowlands claims that an argument can be made for the extension of cognitive processes *only if cognitive processes are specified non-intentionally*. But this means that there is no reason to see the extension of cognitive processes as concerning the extension of the mental rather than the extension of the enablers of mental states.

EMT vs. Content Externalism

Like Clark and Chalmers, Rowlands distinguishes his version of EMT from content externalism. He claims that EMT is more radical and broader in scope than content externalism and distinguishes these externalist positions by contrasting each with internalism. Here I examine how Rowlands makes this distinction in order to clarify his claims with regard to his version of EMT and I begin with a brief explication of how he characterises internalism.

> [Internalism] is both widespread and tenacious, not only as an explicit doctrine but, more significantly, in the clandestine influence it has on explicit doctrines of mind. The philosophical thesis from which the view is born is spelled out by Descartes, and its association with him is sufficiently robust for it to be called the *Cartesian conception*. (Rowlands 1999: 3)

This Cartesian conception, says Rowlands, sees mind as substance; as an object that thinks and which, due to its capacity for thought, must be fundamentally different from other objects, which are incapable of thought. This feature of the Cartesian view of mind has been accused of promoting the myth of the *ghost in the machine* (Ryle 1949). Importantly, although there has been a revolt against Cartesian views of the ontology of mind, most have limited their rejection of Cartesian thinking to a rejection of its dualist assumptions; i.e. a rejection of the 'ghostly' element of the myth. A key part of Descartes' theory that is not generally challenged (as is seen by the commitments of modern day Cartesian materialists) is the idea that the mind is somehow *in* the machine[14] It is this aspect of Cartesianism—internalism—has become ingrained in much contemporary thinking about the mind. Consequently, though there are very few remaining dualists today, the majority of materialists still cling to this facet of Cartesianism and are internalists (Rowlands 1999: 4).

Rowlands claims that this Cartesian internalism has yielded three broad strands comprising internalist ontological, epistemological, and ethical commitments (2003b). My focus is on the ontological and epistemological commitments. According to Rowlands, the ontological commitments of internalism comprise two theses: one concerning the location of mental phenomena and the other concerning possession conditions for mental phenomena by subjects.

> *The Location Claim*: any mental phenomenon is spatially located inside the boundaries of the subject, S, that has or undergoes it.

[14] There are two reasons for this. Many still think that contents cause behaviour and also that causes must be local – hence they must be in the head/body. It is this causalist commitment that really sponsors modern day internalism.

> *The Possession Claim*: the possession of any mental phenomenon by a subject S does not depend on any feature that is external to the boundaries of S. (Ibid: 13)

As Rowlands makes clear, the principal distinction between EMT and content externalism is that content externalism only challenges the Possession Claim, whereas EMT challenges both the Location and Possession Claims. It is for this reason that Rowlands holds EMT to be more radical than content externalism.

Rowlands also sees EMT as being broader in scope than content externalism. He argues that content externalism is fairly restricted in what it applies to, specifically, mental states for which semantic content plays a role in determining the type of mental state they are: "states that possess semantic content essentially" (Ibid.: 134). For this reason he argues that content externalism does *not* apply to cognitive processes or cognitive architectures, nor does it apply to sensations or perceptions unless phenomenal content can be reduced to semantic content (see Rowlands 2003b: 125-134). Although this is true we need not concern ourselves with the scope of content externalism since we are more concerned with understanding EMT, and in this regard the relationship of content externalism with the Location Claim is what matters.

The Location Claim, as concerned with the location of mental phenomena, applies most naturally to mental particulars, where mental particulars are defined as: "concrete, non-repeatable, event-, state- and process-*tokens*" (Ibid: 13) such as Inga's conscious entertainment at time t of the belief *that the museum of Modern Art is Located on 53rd street*. EMT is said to be more radical than content externalism because it challenges this Location Claim, whereas a rejection of the Possession

Claim in the form of content externalism is perfectly compatible with an endorsement of the Location Claim according to Rowlands.

The Possession Claim is about how, in general, mental phenomena are individuated and, as such, it applies most naturally to mental properties, where mental properties are defined as: "abstract, multiply-exemplifiable, event-, state- and process-*types*" (Ibid.). According to the Possession Claim mental properties are non-relational in the important sense that the possession conditions for having a mental property do not depend on anything external to subjects. However, what counts as external to a subject depends on how you define a subject. A typical position for a Cartesian materialist to adopt is to identify the subject with the brain or body (see Rowlands 2003b: 20-25). On this view of a subject, the Possession Claim has it that the possession of a mental state by a subject will not depend on anything that is external to the brain or body; it is determined solely by the physical state of the body or brain (Ibid: 16).

There is an initial difficulty in apprehending such a claim, however, as it seems to be obvious that the mental life of a subject can be dependent on factors that are external to the body. For example, the death of loved one causes me to be sad, and when Inga hears from a friend of a particular exhibition at the Museum of Modern Art she is prompted to recall the location of the Museum. It seems that there is a causal relation between my mental properties (internal to me on this view) and environmental factors (external to me). Rowlands says that the Possession Claim need not and does not deny that this can be the case because the Possession Claim is a claim about the *individuation* of mental properties rather than a claim about the *causation* of mental properties.

For example, Inga believes *that she is walking to the Museum of Modern Art*. This could be because she is actually walking to the Museum of

Modern Art or it could be because she is the victim of a Matrix-style computer-generated simulation. According to the Possession Claim, Inga has the same belief in each case because her internal physical states are identical in each case—external factors are irrelevant to her having that particular belief or beliefs in general. It is consistent with the Possession Claim that external factors can and do have a *causal* influence on Inga's mental states but this influence can only be instrumental, which is to say that external factors can influence mental states only to the extent that they can bring about certain internal physical states.

> And it is the presence of these internal states, *however these states get to be caused*, that determines which mental properties a subject possesses or exemplifies. (Ibid: 17)

So, a full causal story of how Inga comes to acquire her belief in each case will involve external items but these external items have no bearing on the possibility of individuating Inga's belief, they are incidental to Inga's believing *that p*. On this view, Inga's possession, or instantiation, of the belief *that she is walking to the Museum of Modern Art* depends only on features that are internal to her.

Rowlands argues that content externalism challenges the internalist Possession Claim but not the Location Claim. The content externalist challenge to the Possession Claim seeks to establish that certain types of mental states are individuation-dependent on environmental factors but, says Rowlands, establishing this does not entail that a mental state must be externally located. Thus, even if true, content externalism, by itself is not a threat to the Location Claim

> Even in the case of propositional attitudes…the arguments for content externalism show only that such attitudes are externally *individuated*,

> not that they are externally *located*. That is, while the arguments for content externalism entail rejection of the Possession Claim that is partially definitive of internalism, these arguments, by themselves, are not sufficient to entail rejection of the Location Claim." (Ibid: 135).

To see why the individuation dependence of a mental state on environmental factors does not entail the external location of that mental state Rowlands invites us to consider the following analogy he takes from McGinn. According to McGinn the property of being a planet looks to be an externally individuated property because:

(i) specification of which planet something is requires reference to the star that it orbits;
(ii) one cannot know which planet something is unless one knows which star it orbits
(iii) it is not possible for something to be a planet unless it orbits a star;
(iv) one could not master the concept of a planet unless one had also mastered the concept of a star (and the concept of an orbit). (Ibid: 136)

If this is correct and the property of being a planet is individuation dependent, at least in part, on external factors then the question is whether or not this has any entailment with regard to location. Clearly in this case there are implications for location built into to the property of being a particular planet since it requires being a body that orbits a particular star. But Rowlands' point seems to be that the individuation dependence of the planet on the star does not carry entailments with regard to the metaphysical extent of the planet, i.e. the constitutive parts of the planet do not extend:

> [E]ven though the property of being a planet is externally individuated, this does not mean that an instance of this property – an individual planet is located, even partly, where its star and orbit are located. (Ibid.)

We can question whether or not this is the best analogy for Rowlands to make his point. But we can allow the conclusion for the purposes of argument. In general, his claim is that the fact that the property of being x is individuated by the property of being y does not entail that x is located where y is.

> In a similar vein, it would seem that mental properties – for example, the property of being propositional attitude *P* – can be individuation dependent on environmental objects and properties, and so can be externally individuated, without instances of that property – individual propositional attitudes of this type – being located, even in part, where those environmental objects and properties are located. The arguments for content externalism, that is, seem to establish only that mental properties are externally individuated. They do not establish that individual propositional attitudes – instances of those properties – are externally located. (Ibid: 136-137)

If this is correct then perhaps it was misleading of Putnam to say "meaning just ain't in the head". But more importantly for our purposes if the above is correct then arguments for content externalism challenge the Possession Claim without challenging the Location Claim. What allegedly makes EMT interesting is that it seeks to go further than content externalism by challenging the Location Claim. According to Rowlands, some mental phenomena are at least partially located outside of the body of the subject of those mental phenomena.

We looked briefly in chapter one at how Clark and Chalmers saw their version of EMT, Active Externalism, as differing from content

externalism. They characterised content externalism as recruiting passive and distal external features situated at the end of a long causal chain that was not relevant in the here-and-now production of behaviour. By contrast, they saw their Active Externalism as playing an ineliminable role in the production of behaviour and Clark later comments that it is "more like an environmentally extended case of narrow content than a case of broad content" (Clark, forthcoming).

We can use this distinction between narrow and broad content to help us better appreciate what exactly is at issue for EMT theorists such as Clark and Rowlands. Broad content is necessarily fixed, in part, by environmental factors. It is the content attributed to Inga and Twin-Inga in virtue of which we can type-distinguish their respective thoughts *that water (H20) is wet* and *that water (XYZ) is wet*. However, since Inga and Twin-Inga are physically identical there is a sense in which they can be said to be in the same mental state. Certainly, their behaviour will be physically indistinguishable. On these grounds it has been argued that there must be a narrow mental state that causes such behaviour and that if it causes rational, intelligent behaviour then it must be a contentful mental state. Such mental states must have narrow content. Narrow content, if it exists, is the kind of content that supervenes solely on the subject's internal physical states, the kind of physical states that are causally relevant in the production of behaviour. According to Fodor, an original defender of this idea, broad content is a function of narrow content as relativized to environmental context, thus the latter kind of content can vary for physically identical twins like Inga and Twin-Inga when they inhabit different environments.

Accordingly, causally efficacious yet contentful physical states remain internal and a common strategy for internalists is to accept the arguments of the content externalist but seek to limit their force by insisting that

broad content is not the only kind of content. It is exactly this internalist move that EMT theorists want to challenge. Proponents of EMT such as Rowlands seek to undermine this picture of narrow mental states by arguing that physical states that are causally efficacious in the here-and-now production of behaviour are not necessarily located within the body of the subject but can extend into the world.

But, as we will see later on, because EMT accepts too much of this internalist picture that it seeks to challenge, the reduction of EM-style arguments for the partial external constitution of mental states to arguments for the extension of internal mental states with external enablers is always a live possibility: arguments for EMT can be converted into arguments for EMMT.

To see the intended target of the EM theorist clearly it is useful to consider Rowlands' distinction between the task of psychotectonics and the task of psychosemantics. Psychotectonics is essentially an engineering question: 'how can one build a mind?' (Rowlands 1999: 2). Psychosemantics, by contrast, seeks to give an account of mental representations *qua* representations—as states that have content or possess semantic properties. The content externalist arguments that we looked at above attempt to address the questions of psychosemantics, arguing that a person cannot be in possession of a certain type of mental property, or cannot be in a certain type of mental state, unless they stand in the appropriate relation to some object or property in the environment. Because of this such arguments are unconcerned with the project of psychotectonics and thus are ineffective against the internalist Location Claim. Rowlands' Environmentalism and Vehicle Externalism are, by contrast, psychotectonic projects. He is explicit about this:

> [L]et me make it clear that the environmentalist position to be defended in this book claims that cognitive processes are not just externally individuated, they are also, and perhaps much more importantly, externally located. (Ibid: 32)

> Vehicle externalism entails rejection not only of the internalist Possession Claim, but also of the Location Claim. Many mental phenomena, if vehicle externalism is true, are not merely *individuation dependent* on what is going on in the environments of their subjects, but they are also, in part, *located* in those environments. (Rowlands 2003b: 182)

What we need to consider as we examine Rowlands' arguments in support of these claims is whether they entitle him to conclusions of this sort or whether, like Clark and Chalmers, a reading of his version of EMT only entitles him to make claims about the extension of cognition-enabling and belief-forming mechanisms. I begin this examination by looking at Rowlands' arguments for the extension of cognitive processes into the world.

The Ontological Claim

In *The Body in Mind* (1999) Rowlands defends his Environmentalism, which he expresses as a conjunction of two claims:

> *The Ontological Claim*: Cognitive processes are not located exclusively inside the skin of cognizing organisms.

> *The Epistemological Claim*: It is not possible to understand the nature of cognitive processes by focusing exclusively on what is occurring inside the skin of cognizing organisms. (Ibid: 22)

Rowlands says that the Epistemological claim is a corollary of the Ontological claim—if cognitive processes are not located exclusively within the skin of the cognizing organism (The Ontological Claim) then one cannot understand these processes by focusing exclusively on what goes on within this boundary (The Epistemological Claim). Thus, for Rowlands, a defence of Environmentalism consists primarily in a defence of the Ontological Claim; if the Ontological Claim can be established then the Epistemological Claim follows according to Rowlands. Thus, my focus is on this Ontological Claim. Now, it is important to note that Rowlands does not argue that *all* cognitive processes extend into the environment in the manner suggested by the phrasing of the Ontological Claim. Rather, the claim is that certain cognitive processes do extend, while others may be exclusively internal processes (see also Clark 2008b). Furthermore, Rowlands is not claiming that the cognitive processes that do extend into the environment are completely external. Rather, the claim is that these cognitive processes are hybrid entities, straddling both internal and external processes.

The Ontological Claim may well present a direct challenge to an internalist location claim made about cognitive process. But the Ontological Claim challenges the Location Claim specified above in terms of mental phenomena only if a cognitive process counts as a genuine mental phenomenon. Later, we will see that Rowlands' requirement that cognitive processes be specified non-intentionally opens the door for opponents to reject cognitive processes as genuinely mental phenomena.

But regardless of the possible failure of the Ontological Claim to challenge the Location Claim, there is a possible obstacle to progress regarding an argument for the extension of cognition that is commented on by Rowlands and, as we will see in chapters four and five, it is no

small obstacle. It is, in fact, one of the primary hurdles in the path of the proponent of EMT and is one that I flagged in the last chapter. It is the lack of any clear definition of what a cognitive process actually is. This also holds with regard to mental states. And if EMT proper requires the extension of mental states then a mark of the mental is required if the EMT project is to succeed.[15] Rowlands concedes that there is much uncertainty surrounding the concept of a cognitive process and that, consequently, a clear definition of 'cognitive process' that is agreeable to all is hard to come by (1999). Nevertheless, he argues that clarifying the notion of a cognitive task is central to explaining the concept of a cognitive process and that we can approach a definition of a cognitive process via the definition of a cognitive task (Rowlands 2003b: 161). Thus, Rowlands defines a cognitive process as follows:

> A process P is a cognitive process if and only if (i) P is essential to the accomplishing of a cognitive task T, and (ii) P involves operations on information bearing structures, where information carried by such structures is relevant to task T. (Rowlands 1999: 103)

But what is a cognitive task? Rowlands states that a cognitive task is defined partly by way of ostension. That is, we define a cognitive task by pointing to particular examples of cognitive tasks: the task of perceiving the surrounding environment, the task of committing information to, and retrieving it from, memory, or the task of reasoning, for example. So if, for example, I am faced with the task of recalling the location of the museum of modern art and initiate some process that involves operations on information-bearing structures that are essential to the completion of this task, then this process is a cognitive process according to Rowlands. He argues that in certain cases these cognitive processes that allow us to

[15] Rowlands is one of the few EM theorists to take up this challenge. He provides his version of a well-thought out mark of the cognitive in a 2009 paper. We will examine this in a later chapter.

accomplish cognitive tasks are made up of internal operations along with the manipulation of certain environmental structures.

The Manipulation Thesis

Rowlands' arguments for the extension of cognitive processes into the environment are defended in *The Body in Mind* by appeal to his manipulation thesis, where manipulation is understood in what he calls a broad sense, which is to say that manipulation does not necessarily mean manual interaction with the environment. Furthermore, manipulation in this case need not be intrusive—it need not change the environment. What Rowlands' broad conception of manipulation, properly understood, encompasses is any form of bodily interaction with the environment that makes use of the environment in order to accomplish a particular task (Ibid: 23). A prime example of this broad sense of manipulation is that of a sponge making use of the ambient currents by opening and closing certain flagella to maximize the flow of water through it. The interaction in this case is not intrusive, says Rowlands, no aspect of the environment need be changed but the sponge is making use of the environment to accomplish a particular task—the task of maintaining the flow of water through the sponge in order to feed (Ibid.).[16] Applied to cognition, the manipulation thesis yields the central claim of Rowlands' *Environmentalism*; that certain cognitive processes include the manipulation, as just described, of environmental structures.

But it is not *just any* manipulation, so understood, of environmental structures that helps constitute cognitive processing on Rowlands' view. It is essential to his account that the environmental structures being manipulated are information-bearing. In the case of the sponge

[16] We can question, perhaps, whether the flow of water is changed, perhaps in its speed or direction, because it flows through the sponge but the point is not a crucial one.

manipulating the ambient currents in its environment, no information-bearing structures are being manipulated, thus, even if the sponge's task could be described as cognitive, the manipulation of the ambient currents could not be described as part of a cognitive process because it does not involve the manipulation of information-bearing structures.[17]

Cognitive processes, on this account, must involve the manipulation of information-bearing structures for the purpose of the completion of a certain cognitive task. And if these information-bearing structures should happen to be located beyond the boundaries of skin and skull then the cognitive process also extends beyond the boundaries of skin and skull. So, if a process that is essential to the completion of a certain cognitive task involves the manipulation of some external (environmental) information-bearing structure then this process is cognitive and extends into the environment.

We can run this argument from the perspective of the Parity Principle endorsed by Clark that we examined in chapter one. If a cognitive process is a process that facilitates the completion of a particular cognitive task and if this process proceeds by way of processing certain information that is relevant to the completion of the task then it should not matter where this process is carried out in so far as its cognitive status is concerned. What matters is that the putative cognitive process fulfils the criteria laid out by Rowlands above—which is to say that the process

[17] We can see how Rowlands' manipulation thesis as applied to cognitive processes differs from Clark and Chalmers' use of epistemic actions. Whereas epistemic actions, as defined by Kirsch and Maglio and employed by Clark and Chalmers, make some change to the environment that aids and augments cognition, the manipulation thesis does not hold that every instance of manipulation yields a change to the environment. We could see, perhaps, epistemic actions as special cases of manipulation as defined by the manipulation thesis. On this view all epistemic actions exemplify the manipulation thesis as applied to cognitive processes but not all cases of the manipulation of information-bearing structures performed for the completion of a cognitive task are epistemic actions.

is essential to the completion of a cognitive task and involves processing on information-bearing structures.

Rowlands' Environmentalism as noted briefly in chapter one, endorses the correct reading of the Parity Principle in that it anticipates the arguments of the cognitive integrationist and emphasises the differences that exist between the external portion of extended cognitive processes and their internal portion. The parity that Rowlands sees as existing between the purely internal cognitive process and the extended cognitive process exists on the level at which cognitive status is awarded, not on the level of what implements the processing. Rowlands recognises that there are important differences that exist between purely internal and extended aspects of cognitive process and, much like the cognitive integrationist, Rowlands argues that these differences are required if the external portion of extended cognitive processes are to be in any way useful in assisting the completion of specific cognitive tasks.

To see why this should be the case I will examine some arguments from the evolutionary perspective that Rowlands develops in support of Environmentalism. The first element of the evolutionary argument that I want to consider is Rowlands' *Barking Dog Principle*.

The Barking Dog & Evolutionary Support for EMT

> *The Barking Dog Principle*: If it is necessary for an organism to be able to perform a given adaptive task T, then it is differentially selectively *dis*advantageous for that organism to develop internal mechanisms sufficient for the performance of T when it is possible for the organism to perform T by way of a combination of internal mechanisms and manipulation of the external environment. (Ibid: 80)

In order to demonstrate why such a principle should hold, Rowlands invites us to consider the concept of evolutionary cost. Any evolutionary development requires an investment of resources. Each organism has a finite amount of resources, such that an investment of resources—in the development of wings, for example—will mean fewer resources to invest in other areas, such as reproduction. Every evolutionary adaptation costs something and costs typically multiply. A larger brain, for example, means a larger head, which means more weight at the front of the body, which means developing a longer tail to balance this extra weight and so on. So, at every stage of evolutionary development there is a balancing of costs and benefits.

Rowlands' Barking Dog Principle appreciates the evolutionary cost involved with adopting particular evolutionary strategies and recommends the adoption of the manipulate-the-environment strategy because it can be purchased at less evolutionary cost. Why do all of the work if you can get someone, or something to do some of that work for you? Or, if you have a dog then you do not have to bark yourself (Ibid: 80).

Andy Clark also recognizes this principle, albeit by another name. He claims that if evolution can economize by exploiting the structure of the physical environment in order to aid an animal's processing then it is very likely to do so. He calls this the 007 principle:

> *The 007 Principle*: In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done. (Clark 1989: 64)

So it makes good evolutionary sense, according to Clark and Rowlands, to delegate certain tasks, or at least a significant portion of certain tasks, to the external environment where possible.[18] The reason that this strategy of delegating certain tasks to external environment makes sense according to Rowlands is that it enables an organism to maximize differential fitness. He illustrates this quite nicely in discussion of the evolutionary strategy adopted by the beaver. Beavers build dams because a dam results in the creation of a miniature lake, which means the beaver can travel greater distances by water, thus making the task of reaching and transporting food much safer. The building of dams is, in this way, advantageous to the beaver so it became part of the beaver's evolutionary strategy. However, the beaver could have evolved in other ways. Imagine that in the past the beaver's ancestors began evolving in two distinct ways. One way involved adapting the dam-building strategy and the other strategy concentrated on making the beaver stronger, quicker, more intelligent etc.

The tasks that the beaver's ancestor had to accomplish were the location and transportation of food and the evasion of predators. The conjunction of these two tasks, Rowlands calls 'T'. Now, let us suppose that S(o) is the strategy adopted by the dam-building or ordinary beaver and S(s) is the strategy adopted by the other, super-beaver. If both the ordinary beaver and super-beaver are equally competent in performing T then it follows that the benefits for the ordinary beaver through adopting S(o) are equal to the benefits for the super-beaver through adopting S(s). This does not entail, however, that the differential fitness of each beaver is the same. This would only be the case if the cost of adopting S(o) was the

[18] It should be noted at this point—in anticipation of what is to follow in later chapters—that even if it makes evolutionary sense for an organism to delegate cognitive tasks to the environment this does not necessarily entail that the processing taking place in the environment is *cognitive* processing. The only thing that matters in this case is whether or not this processing exemplifies the mark of the cognitive.

same as the cost of adopting S(s) and this is said not to be the case. Rowlands claims that S(s) is bought at a greater cost than S(o). The implementational cost of S(o) comprises the genetic resources needed for development and maintenance of the large flat powerful teeth and tail and the muscle that surrounds each. In the case of S(s) the implementational costs seem greater. The super-beaver must develop more powerful muscles for dragging its food over long overland journeys so its limbs and torso must become bigger. It will have to become quicker to escape predators, which entails longer limbs. It might have to become more intelligent and will also have to improve its senses so as to detect and avoid predators, which would require encephalization, which leads to a larger head, which means more weight at the front of the body, which must be balanced at the rear, and so on.

The disparity in performance cost between the two strategies is even greater. Although the ordinary beaver has the large initial cost of building the dam it seems small compared to the daily overland journeys with risk of predation that the super-beaver must make. So, both the implementational and performance costs of S(o) are less than S(s). With respect to task T (finding food and evading predators) the differential fitness of the ordinary beaver would be greater than that of the super-beaver. Thus, from the point of view of evolutionary fitness it seems to be differentially selectively advantageous to accomplish a particular task by developing the capacity to manipulate the environment instead of accomplishing that same task by developing additional, improved, or enlarged internal mechanisms or structures (Rowlands 1999: 79).

This leads us to another principle that Rowlands highlights and which also evidences his support of the Cognitive Integrationist position. In the case of the beaver consider, once more, the task, T, that faces him. The beaver must locate and transport food while evading predators. Now, the

internal mechanisms that the beaver has, in fact, evolved to accomplish this task are not at all obvious given the nature of the task. This insight yields the *Principle of the non-obvious character of evolved internal mechanisms*. Rowlands states it as follows:

> NOC: For the performance of a given task T, and for any internal mechanism M which has evolved in organism O and which, when combined with suitable environmental manipulation on the part of O, allows O to perform T, the nature of M is not always obvious on the basis of T. (Ibid: 81).

According to Rowlands, this principle applies equally to internal mechanisms that have evolved in conjunction with the manipulation of environmental structures for the purposes of completing specific *cognitive* tasks. These internal mechanisms or structures will by no means be obvious given the nature of the cognitive task to be completed. The reason that these internal mechanisms or structures are not obvious on the basis of the task to be completed is precisely because they have been developed in conjunction with a strategy of the manipulation of environmental structures. Once you examine the bigger picture, incorporating the integrated internal and external portions of the extended process, the integrated mechanism, as a whole will become more obvious given the nature of the cognitive task. It is only upon consideration of the beaver's manipulation of the structures in his environment that his evolutionary strategy of the development of powerful jaws, large teeth and a large flat tail along with the associated musculature makes sense.

Furthermore, it makes sense only when the extended process, mechanism, or strategy as a whole is considered, precisely because the internal portion of the process, mechanism, or strategy is insufficient for the completion of the task in question. It is insufficient, on its own, for the

completion of the task because it requires something different, something extra, to complete the task. This is exactly what the cognitive integrationist is getting at when he talks about the integration of distinct internal and external processes. The precise reason that the external portion of the process is integrated with the internal process to create one extended process is because on its own the internal portion of the process is insufficient for, or incapable of, completion of the task. In this way Rowlands' Environmentalism, and his evolutionary arguments in support of Environmentalism, endorses the insights of the integrationist perspective.

Rowlands may well be correct in arguing that it is a biological tendency for organisms to offload significant portions of tasks onto the environment where possible. He may be correct in arguing that doing so makes good evolutionary sense from the perspective of the differential fitness of the organism. He may be correct that this holds true with regard to cognitive tasks as well as more mundane tasks that do not involve cognition. He may be correct in arguing that offloading portions of cognitive tasks onto the environment transforms the nature of the internal processing that needs to be carried out in order to complete the cognitive task in question. And he may be correct in arguing that, in cases where a portion of the cognitive task is offloaded onto the environment, the manipulation of the information-bearing structure in question is essential to the completion of the cognitive task. None of this, however, entails that cognitive processes extend into the environment. What is required for such a conclusion is that the extended hybrid process, comprising internal and external processing, displays the mark of the cognitive, whatever that turns out to be.

What Rowlands has given us is a good supporting argument for why we might expect organisms to adopt the 'manipulate the environment'

strategy with regard to cognitive processing. What he needs next, is an argument to show that adopting this strategy entails the extension of cognition into the world.

Cognitive Extension

Before I examine Rowlands' arguments for the extension of cognitive processes I want to give a reminder of his definition of a cognitive process, which is key in this section:

> A process P is a cognitive process if and only if (i) P is essential to the accomplishing of a cognitive task T, and (ii) P involves operations on information bearing structures, where information carried by such structures is relevant to task T. (Ibid: 103).

In *The Body in Mind* Rowlands examines perception, memory, reasoning and language acquisition, arguing that Environmentalist accounts can be given in each case that endorse the Ontological Claim. From these examinations we can derive two key claims that Rowlands endorses with regard to cognition.[19]

1. The amount of internal information processing that an organism needs to perform in order to accomplish a cognitive task is inversely proportional to the amount of relevant information that is available to that organism in the environment.

[19] There are actually four general claims made by Rowlands in *The Body in Mind* regarding cognitive processes. However, one claim—that we cannot begin to understand the internal information-processing task facing an organism for the completion of a cognitive task unless we understand what relevant information is contained in the organism's environment—relates to the Epistemological Claim rather than the Ontological Claim, which is not of concern here. And another is simply the claim that in certain cases manipulating external structures is a form of information processing. Rowlands admits that he does not require this claim in order to run his argument so I propose to ignore it here.

2. An organism can process information relevant to the completion of a cognitive task through acting upon, and thus effecting transformations in, information-bearing structures in the environment.

If we allow these two claims to pass unchallenged (which I propose to do here for the sake of argument) and if the definition of a cognitive process proposed by Rowlands above is correct, then cognition extends into the world. I now examine these general claims applied to specific cognitive processes, beginning with perception.

In making his case for the extension of the cognitive process of visual perception into the world Rowlands examines, first of all, a typical internalist account of perception (in the style of David Marr) in order to contrast with his own. Such an account, he claims, is typically organised around the following sort of framework:

1. Perception begins with stimulation by light energy impinging on it.
2. This results in a retinal image, characterised in terms of intensity values distributed over a large array of locations.
3. Retinal images carry relatively little information, certainly not enough to add up to genuine perception.
4. In order for perception to occur, the information contained in the retinal image has to be supplemented and embellished (i.e. processed) by various information processing operations.
5. These information-processing operations occur inside the skin of the perceiving organism. (Rowlands 2003b: 169-170).

So, this sort of framework is characterised by the positing of a retinal image carrying insufficient information. Given the paucity of information contained in the retinal image, it must be processed by a number of internal operations in order for visual perception to be possible. So,

according to Rowlands, the retinal image is always the starting point for analysis of visual perception according to this internalist framework and because the retinal image is inside the skin of the cognizing organism, the processing that is carried out on the retinal image is also seen as occurring inside the organism's skin. So, on this view visual perception is naturally construed as a purely internal process. Rowlands turns to the work of J.J. Gibson for an alternative to this internalist account.

For Gibson, the starting point for theorising about visual perception is not the retinal image but the optic array. The optic array is the structure of light that surrounds an observer yielding information about the environment, information which changes as the observer moves because the movement alters the structure of the optic array. Thus, the optic array is, says Rowlands, an external information-bearing structure. Given the existence of the optic array, Rowlands makes the following claim which is a specific application of the more general Claim 1 above:

> The amount of internal information processing which an organism needs to perform in order to accomplish visual perception task T is inversely proportional to the amount of relevant information that is available to that organism in the optic array. (Rowlands 1999: 108)

This claim follows once we allow that the optic array exists, which, Rowlands states, is not controversial. What this claim makes explicit, says Rowlands, is that perception is essentially bound up with action. The perceiving organism moves in order to change the structure of the optic array that surrounds it and thereby acquires the information contained within. This yields the following claim, which is a specific application of the more general Claim 2 above:

> An organism can process information relevant to visual perception task T through acting upon, and thus effecting transformations in, the optic array. (Ibid: 116)

If this is correct then at least some of the information processing that is essential to the completion of the cognitive task of visual perception lies outside of the boundaries of skin and skull. Rowlands says that if the manipulation of the external optic array is a process that is essential to the completion of the cognitive task of visual perception and if the definition of a cognitive process above is correct then at least some of the processing involved in visual perception is external, which is the message of the Ontological Claim. Thus, the Ontological Claim can be defended with respect the cognitive task of visual perception according to Rowlands.

A similar result follows if we look at Rowlands' arguments for the extension of memory. We all use external cues to aid memory. I have a diary, for example, where I write down everything that I need to do in the course of the day because I don't trust myself to remember 'on my own.' Examples of this type of memory aid abound: the knot in the hanky, the shopping list, asking somebody else to remind you of something, or using a 'wearable remembrance agent'[20], are all possible ways of using external cues that trigger the process of remembering. The standard reading of such a situation is that I see the knot in the hanky, which triggers something within my brain that causes me to remember whatever it is that I wanted to remember. The remembering is typically seen as a purely internal process and the memory aid as external to the real magic of memory. Memory cues or memory aids are seen as just precisely that: something to kick-start the real process of remembering. This is not

[20] Apparently it exists, called a "Private Eye" and made by Reflection Technology. For discussion see: Clark 2003: 46.

necessarily the correct way of looking at things according to Rowlands. He claims, at least with regard to the memory systems possessed by modern literate adult human beings, that there is no sound theoretical reason for imposing such a dichotomy between internal memory processes and external 'aids' to those processes (Ibid: 121).

Rowlands begins by considering episodic memory, which is memory of specific episodes in the life of the subject. Although modern literate adults do possess an episodic memory system, it does not seem to operate with anything like efficiency or accuracy of people of preliterate cultures (Luria & Vygotsky 1992). Similarly, children rely much more heavily on episodic memory than adults (up to a certain point in their development). If the episodic memory of preliterate cultures and children is superior to that of modern, literate adults, then it seems legitimate to wonder if the memory in general of preliterates and children could be said to be superior to that of modern, literate adults. Rowlands argues that this is an incorrect way of looking at the situation, stating that the difference here is not one of degrees of superiority but, rather, one of strategy. Modern literate adults simply employ a different memory storage strategy; they employ semantic memory, which is, roughly, memory of facts—remembering *that* something is the case. Modern literate humans do possess episodic memory systems but this has been superseded by semantic memory says Rowlands (1999). He claims that this difference in memory strategy exists despite the lack of any relevant biological difference between preliterate people and modern literate humans, or between children and adults, at least none that could explain the switch in storage strategy from episodic to semantic.

Rowlands says that the explanation of the development of this alternative strategy lies in a change in the functional hardware involved in memory storage and retrieval. But he argues that this change in hardware does not

have a biological explanation in the human brain. Instead, Rowlands argues that the development that made this alternative strategy possible was a fundamental change in the environment in which the modern brain is situated; the development of external means of representation (Ibid.).

Given the availability of an external means of representation how would we imagine a memory storage and retrieval strategy to develop? Rowlands considers the use of a system of knots for communication called *kvinus* (Rowlands 1999: 134-137; 2003b: 178-177). Kvinus were used in Peru to record chronicles, to transmit instructions to remote provinces, and to record information about the state of the army among other things. A kvinu officer was appointed whose task was the tying and interpreting of the knots such that the system became standard and could be used to record all major matters of the state. Rowlands compares the strategy of the Peruvian kvinu officer with the strategy of the preliterate African envoy, both charged with the task of transporting a message. The African envoy (who cannot make use of external means of representation) will need to remember the content of the message word for word, in the correct order. However, the Peruvian kvinu officer need not remember either the content or order of the message, all he needs to remember is the code that will allow him to extract the information that is contained in the knots. So the African envoy will need to employ his episodic memory each time that he needs to remember a new message. By contrast, the Peruvian kvinu officer need only employ his episodic memory once in learning the code. So, given the availability of an external means of representation in the form of the kvinus, we can say that the more information that is carried and processed externally, the less information needs to be carried and processed internally. Thus, we get a specific application of the more general Claim 1 offered above:

> The amount of internal information processing that an organism needs to perform in order to accomplish memory task T is inversely proportional to the amount of relevant information that is available to the organism in the physical structures around it. (Rowlands 1999: 122)

Now although the kvinu officer must employ some internal processing in order to remember the code that allows him to translate the knots, the remainder of the information is carried by the knots themselves and by using or processing this information the kvinu officer is able to complete his particular memory task. So, the kvinu knots are structures that carry information, that is, they are information-bearing structures which the kvinu officer must manipulate. Furthermore, the manipulation of these information-bearing structures is essential to the completion of the cognitive task of retrieving, or remembering, the message. Thus, we get a specific application to memory of the more general Claim 2 above:

> An organism can process information relevant to memory task T through the manipulation of physical structures in its environment. (Ibid.)

If Rowlands' definition of a cognitive process is correct then the Kvinu officer's use of the system of knots to relay messages forms part of a cognitive process. In this case the information-bearing structure to be manipulated happens to be located beyond the boundary of skin and skull and so, says Rowlands, the cognitive process in this case also extends beyond this boundary. Thus, the cognitive process of remembering can extend into the environment in accordance with the Ontological Claim.

Rowlands also considers the possibility of the extension of reasoning via consideration of a classic example of mathematical reasoning that we examined briefly in chapter one, which by extension, he claims, provides

a template for the operation of formal reasoning in general. There is no need to rehearse the long-multiplication example that we examined in chapter one. Rowlands claims that the numbers written down on the piece of paper constitute an information-bearing structure and these structures are manipulated for the process of completing the long multiplication task, which is a specific application of the more general Claim 2 above. Furthermore, without adopting this strategy the possibility of completing the task diminishes drastically unless one is particularly adept at mental arithmetic. So, the amount of internal processing that a subject must perform in order to complete the task is inversely proportional to the amount of relevant information in the subject's environment, which is a specific application of the general Claim 1 above. Thus, if Rowlands' definition of a cognitive process is correct, then reasoning can extend into the environment by being partially constituted by these external information-bearing structures, which provides further endorsement of the Ontological Claim.

We can, says Rowlands, apply these arguments to spatial as well as formal reasoning tasks such as the long multiplication example considered above. Consider how difficult the completion of a jigsaw would be if it was forbidden to pick up the pieces and physically rotate them. Much like the Tetris example considered in chapter one, in this case the external rotation of the pieces does work that would, otherwise, have to be carried out internally. That is, one would have to form a mental image of the piece and mentally rotate it in order to determine its best fit. The fact that we can physically rotate the pieces removes the need for performing these difficult mental imagery tasks. Thus, once more, the amount of internal processing that is required for the completion of the reasoning task is inversely proportional to the amount of relevant information contained in the environment. And, furthermore, the manipulation of pieces in the manner just described constitutes the

manipulation of external information-bearing structures for the completion of the spatial reasoning task—completing the jigsaw. So, again *if* Rowlands' definition of a cognitive process is correct then the process just described counts as cognitive and provides further endorsement of the Ontological Claim.

In later work, Rowlands attempts to broaden the scope of his version of EMT and does so by endorsing a form of vehicle externalism—a variety of EMT that we will consider in detail the next chapter. Here, I will examine how Rowlands uses vehicle externalism to argue for the extension of mental states, as well as cognition, into the world.

Vehicle Externalism

In *Externalism* (2003b) Rowlands puts a slightly different spin on his version of EMT, arguing for a form of vehicle externalism. The vehicle/content distinction is essentially a distinction between content or meaning and what has or carries that content. The typical manner of characterising vehicles is by exemplifying this distinction in simple sentences. So, for example, in the case of the sentence 'water is wet' we can distinguish the content of the sentence, what is expressed by that sentence—the proposition *that water is wet*—from the sentence itself—the order of symbols and spaces of the written sentence or the phonetic sound of the spoken sentence. This syntactic-phonetic sentence form is the vehicle and it carries the content *that water is wet*.

The distinction is also thought to apply to mental contents, which is to say that in the case of particular mental states, such as the thought *that water is wet*, we can distinguish the content of the mental state from the vehicle that has or carries that content. In this case vehicles are the physical processes and mechanisms, such as neurophysiological

processes, that are the causally explanatory events or processes of particular mental content tokens. The vehicles of mental content occupy a differentially token-explanatory role with regard to the mental contents rather than a type-explanatory role. That is, the mechanisms and processes described can explain particular instances of mental state tokens, they explain the occurrence of a particular belief that *p* at time *t*, but do not explain why mental states of this type obtain at all.

This way of using the term 'vehicle' is, as demonstrated, quite tightly linked with the notion of semantic content. But Rowlands argues that there is another way of using the term vehicle that 'bears no direct or essential connection to content' (Rowlands 2006: 31). Instead we can characterise a vehicle, says Rowlands, in terms of the distinction between a state and a process.

> The above conception of a vehicle sees it primarily as a state—one that has a token-explanatory role in accounting for why a token content, and thus why a token mental state, should obtain. But there is another sense of vehicle that applies to mental *processes* rather than mental *states*. And its most natural application is to a subclass of mental processes: cognitive processes. (Ibid.)

Rowlands argues that just as the vehicles of mental content are the subpersonal states that explain particular content tokens, the vehicles of cognition are the subpersonal mechanisms on which particular cognitive processes run. He says that we can refer to these vehicles of cognition as cognitive architecture.

> [C]ognitive processes are implemented in certain subpersonal mechanisms possessed by the cognizing organism. We can refer to these as the cognitive *architecture* of the organism. In this sense, the vehicles of cognition consist in cognitive architecture: the subpersonal

> mechanisms that allow cognitive processes to be run. However, by extension the notion of a vehicle can also be extended to the processes themselves—as long as we are clear that these are identified *non- intentionally*. (Ibid: 32, emphasis mine).

In talking about the vehicles of mental content we can distinguish between the content and the vehicle as bearer of that content. But the question arises as to whether it is legitimate to make a similar distinction in talking about the vehicles of cognition. That is, can we make a distinction between cognition itself and the vehicle of cognition? Rowlands thinks that we can but he believes that the extension of the vehicles of cognition means that cognition extends, provided that we identify the cognitive process non-intentionally.

This admission seems detrimental to the prospects of a genuine EMT. If cognitive processes must be specified non-intentionally then, if intentionality is crucial for mentality, we have no compelling reason to see the extension of cognitive processes as concerning an extension of the mental in any interesting sense.[21] If this is correct then the best that can be decisively concluded from these arguments is the extension of the belief-forming mechanism, the enablers of mental states.

Rowlands does not himself argue for the extension of mental states into the world but in *Externalism* (2003b) he reports on Clark and Chalmers' argument for the extension of 'cognitive states'. It is not immediately clear why Rowlands chooses to specify Otto's belief as a cognitive state rather than a mental state. But later we learn that he sees cognitive states as being the "products" of cognitive processes (Ibid: 182). So, on Rowlands' account, Otto's believing *that the Museum of Modern Art is*

[21] This concerns the mark of the mental which will be discussed in detail in later chapters

located on 53rd Street is a cognitive state that is the result of an extended cognitive process involving the manipulation of the notebook. In this case, even if we accept that this constitutes an example of a genuinely extended cognitive process the option is still there to reject the claim that the mental (or cognitive) state extends. And if the extension of mental states is required for the extension of mind, the extension of cognitive processes that are specified non-intentionally will not entail a genuine EMT.

Constitution vs. Enabling

There are problems with using the notion of a vehicle in the manner just described in reference to cognitive processes. As mentioned, the notion of a vehicle is very tightly bound up with the notion of semantic content and it is not clear that it can be appropriated to be synonymous with 'cognitive architecture' as Rowlands does. Furthermore, it is not clear that the extension of cognitive architecture necessarily entails the extension of cognition as argued by Rowlands. One of the more important corollaries of endorsing the vehicle/content distinction is the warning to be careful not to conflate the properties of vehicles with the properties represented in the contents of mental states (see Hurley 1998b). Thus, it is a mistake to assume that the properties of the vehicle of content will project onto the content itself, and vice versa. For example, it is, quite obviously, a mistake to assume that the vehicle that carries the content 'green' must itself be green.

Something similar *may* hold with regard to the locational properties of cognitive architectures and cognition. If so, the extension of the vehicles of cognition, cognitive architecture, will not entail the extension of cognition. Instead, the extension of the vehicles of cognition may best be described as extending the enabling mechanisms of cognition. In the next

chapter, we will delve deeper into issues surrounding the use of vehicles by the EMT theorist in consideration of the work of Susan Hurley.

However, even apart from the problems Rowlands faces over his use of vehicles there are problems with his account of the extension of cognitive processes as described earlier. Rowlands' arguments for evolutionary support for the manipulate-the-environment strategy may well be correct. However, the success of these arguments demonstrates, at best, that it makes good evolutionary sense to make use of external information-bearing structures in order to aid in the completion of cognitive tasks. In order to demonstrate that cognitive processes extend Rowlands needs to demonstrate not only that adopting this strategy is relevant or essential to the completion of the cognitive task in question, but that the manipulation of the external information-bearing structure constitutes cognitive processing.

So, although Rowlands' suggests that his arguments in *The Body in Mind* entail the extension of cognitive processes into the environment, it is not conclusively demonstrated that he is entitled to this conclusion. One could take Rowlands' arguments as providing a good case for certain information-bearing structures in the environment enabling the completion of certain cognitive tasks without partially constituting the cognitive process in question. Indeed, Rowlands notes that his arguments in *The Body in Mind* should be viewed as providing an alternative picture to the one provided by Internalism, however, and not as conclusively refuting it, and thus it is not a decisive argument against internalism. It seems, then, equally valid is to view processes that involve the manipulate-the-environment strategy as enabling, rather than constituting cognition.

But an opponent of EMT can accept the extension of cognitive processes in the manner argued for by Rowlands and yet maintain that mental states remain internal. If cognitive processes are specified non-intentionally then it is not clear that their extension into the world concerns an extension of the mental in any interesting sense. Thus, we can question the extent to which Rowlands actually argues for EMT since his focus is on extending cognitive processes that must be specified non-intentionally rather than extending the intentionally specified mental states that would entail a genuine EMT. In the next chapter I examine the work of Susan Hurley, suggesting that a similar conclusion can be reached.

Chapter 3

Introduction

Here, the exposition of EMT concludes with Susan Hurley, who is taken to be one of its most meticulous exponents. However, I will argue that if the extension of mental states is required for a genuine EMT then it is not clear that Hurley advocates an EMT of the same kind as Clark and Chalmers. Hurley's vehicle externalism is the thesis that the vehicles of mental content can extend beyond the boundaries of skin and skull and into the world. I examine the arguments that Hurley offers in favour of just such a conclusion and, as with my treatments of the work of Clark and Rowlands, consider whether these arguments actually yield a conclusion of EMMT—the extension of the machinery of mind—rather than a conclusion of EMT. I begin, however, as in the preceding chapters, with an examination of how Hurley characterises her vehicle externalism by distinguishing it from content externalism.

EMT vs. Content Externalism

In 'Varieties of Externalism' (forthcoming) Hurley offers a taxonomy of externalist positions that yields four varieties. She categorises these varieties, first of all, according to the type of explanation offered by each externalist position; distinguishing 'what'-explanations from 'how'-explanations. Hurley describes 'what'-explanations in the philosophy of mind as accounting for why a mental state has content of a certain *type*; rather than some other content, or qualia of a certain type rather than some other qualia.[22]

[22] Hurley's is not arguing that 'what'-explanations account for mental state types, i.e. why a mental state is a belief rather than a desire. The type distinction applies to the contents of mental states rather than their attitudinal aspects.

> For example, they explain why an intention is an intention to look inside the box on the left, rather than to look inside a different box, or to do something else entirely. Or why an experience is one of how something looks rather than of how it feels or sounds, or is an experience of red rather than of green. (Ibid.)

'How'-explanations in philosophy of mind explain the occurrence of particular tokens of mental states, such as the belief, at time t *that p*, by accounting for the processes or mechanisms that *enable* this particular mental state.

> They explain, e.g., what processes or mechanisms enable a given intention to look inside the box on the left, or a given visual experience of a certain surface as blue. (Ibid.)

Hurley argues that we can look at the distinction between content externalism and her version of EMT (vehicle externalism) in terms of the distinction between 'what'-explanations and 'how'-explanations. On this account a 'what'-explanation is an externalist 'what'-explanation if it invokes external features in accounting for why a mental state has a particular content type, e.g. the content *that water (H_20) is wet* rather than the content *that water (XYZ) is wet*, for example. Similarly, a 'how'-explanation is an externalist 'how'-explanation, like vehicle externalism, if it invokes external features in explaining the tokening of a particular mental state, such as the belief by Otto, at time t, *that the Museum of Modern Art is Located on 53rd street*.

The second category of distinction that Hurley employs in her taxonomy of externalist positions is between different *explananda*—what is to be accounted for by the externalist explanation. Here again there are two possibilities according to Hurley: mental content and phenomenal quality. A 'what'-explanation that seeks to give an account, for example, of why a

particular belief is a belief *that water (H_20) is wet* by appealing to external factors is, by Hurley's taxonomy, a 'what'-content externalist position: the content externalism of Putnam and Burge. A position that seeks to explain phenomenal quality type, e.g. why a particular experience is of red rather than green, by appealing to external factors is a 'what'-quality externalist position.

Similarly, an externalism that offers explanations of particular tokenings of semantically contentful mental states by appealing to external factors ('how'-content externalism), can be distinguished from what Hurley says seems a more radical position: offering explanations of particular tokenings of *phenomenal quality* by appealing to external factors ('how'-quality externalism).

Thus, on Hurley's taxonomy her vehicle externalism (which she sees as being of a kind with the positions endorsed by Clark and Chalmers and Rowlands) is a 'how'-content externalist position. Before moving on to consider how Hurley argues for her vehicle externalism it is important to note that in 'Varieties of Externalism' (forthcoming) Hurley distinguishes different types of externalism as offering *explanations* of different mental phenomena. For example, 'what'-content Externalism explains intentional content type, 'what'-quality externalism explains phenomenal quality type, 'how'-content externalism explains intentional content tokens, and 'how'-quality externalism explains phenomenal quality tokens. But, as we have briefly touched on in the previous chapters, the fact that appeal to external factors may offer the best explanation of the occurrence of a particular content token does not necessarily mean that the content token itself is partially externally constituted or located. It may just be that external factors play an ineliminable role in *enabling* that particular content token.

Hurley's position in this regard is not immediately clear. As we saw above she sees vehicle externalism as being of a kind with the EMT advocated by Clark and Chalmers. But in 'Varieties of Externalism' (forthcoming) she interchanges 'vehicle externalism' with 'enabling externalism'; suggesting that she may endorse EMMT rather than an EMT. What needs to be established, then, in looking at Hurley's work, is what her take on vehicles is. Are vehicles constitutive of mental states or are they merely enablers?

Vehicles

Hurley argues that the vehicles of mental content are events or processes that are described at the *subpersonal* level as opposed to the *personal* level (1998b). The personal level, according to Hurley, is the folk psychological level of explanation that we use every day to understand the behaviour of others, by attributing beliefs, desires, and intentions to them and by thinking of them as acting for reasons. Inga goes to the museum of Modern Art; why does she do this? We can understand her actions by attributing certain beliefs and desires to her. Inga has a desire to see a particular exhibition; she believes that this exhibition is at the Museum of Modern Art, that the Museum is located on 53rd street, and she forms an intention to go there. Hurley says that these attributions of propositional attitudes are made at the personal level.

The subpersonal level of explanation yields causal explanations of these personal level content tokens in neurophysiological or computational terms, describing processes, mechanisms, and the functions of theses mechanisms. The contents of mental states, which we attribute at the personal level, are said by Hurley to be *carried* by these processes, which is why Hurley calls them *vehicles* of content (Ibid: 3).

On Hurley's conception these vehicles of mental content occupy a token-explanatory rather than a type-explanatory role. That is, these mechanisms and processes described at the subpersonal level can explain particular instances of mental-state tokens, "particular thoughts on particular occasions" (Ibid: 19), such as the belief *that p* at time t.

However, although a vehicle can explain the obtaining of a particular content token such as the belief *that p* at time t, it will not explain why mental states of this content type (*that p*) obtain in general. Vehicles of content, in other words, don't tell us why the token mental states carried by the vehicles have the contents that they do. And the properties of vehicles of content, thus conceived, do not fix the contents carried by the vehicles. So, for Hurley, the vehicles of content do not include "all the entities back through time needed to explain why mental states or episodes of a given type or with a given type of content exist at all" (Ibid: 28).

For example, in an account of which content is carried by a particular vehicle, type-explanatory processes, such as evolution, are often appealed to and we saw in chapters one and two how contextual factors can be appealed to in taxonomising mental state types. However, as Hurley explains:

> [V]ehicles of content do not include all this context. Vehicles explain the obtaining of particular mental states and processes, even if the contents of the mental states in question are determined by further type-explanatory factors. (Ibid: 28).

Consequently, says Hurley, we can think of vehicles differentially, which is to say that a vehicle is what is left over when you subtract what is needed to explain the existence of some mental state type from what is

needed to give a full explanation of the obtaining of a particular instance of this mental state type. So, in a full explanation of the obtaining of a particular instance of a mental state type, evolutionary, normative and contextual content-fixing factors may be appealed to but an appeal to what pattern of neurons fired may also be required.

To better see Hurley's point consider the following sentence token written in ink on paper: *the cat is on the mat*. A full explanation of the obtaining of this particular instance of that content type will perhaps require an appeal to the historical processes underlying the evolution of language and writing among other things, along with an appeal to the actual sentence token itself; the particular markings of ink on paper. If vehicles do the *differentially* token-explanatory work then the vehicle in this case is what is left over when you subtract what is needed to explain the existence of the content type: *that the cat is on the mat* from what is needed to give a full explanation of the obtaining of a particular instance of this content type. In this case, we are left with the ink markings on paper, which is the vehicle that carries the content.

Initial considerations suggest, then, that Hurley sees vehicles as occupying an enabling role with regard to mental contents rather than a constitutive role. This is something that will be teased out more fully in the course of the chapter but for now, armed with this definition of a vehicle we can proceed to investigate how Hurley argues for the extension of these vehicles into the environment. And we begin with her attempt to arrive at a satisfactory account of the unity of consciousness at a time.

Hurley's Vehicle Externalism emerges from her project of providing an intelligible account of the unity of consciousness. She criticises traditional approaches to the problem because of their failure to give an adequate account what constitutes unity. In attempting to provide such an account she challenges the assumptions of these traditional approaches and illustrates how, in certain cases, the extension of vehicles into the world may be necessary to account for the unity of consciousness at a time.

Unity of consciousness is said to hold when distinct conscious states are co-conscious. So when consciousness that *p* at *t* and consciousness that *q* at *t* combine in one centre of consciousness to give consciousness at *t* that *p* and *q*, then the states are said to be co-conscious, or unified within one consciousness. Obviously then, it is not sufficient for co-consciousness that two experiences occur simultaneously. For example, if Inga is talking with Otto at time *t*, then we can say that Inga has an experience of perceiving Otto's face at time *t* and Otto has an experience of hearing Inga's voice at time *t*. Clearly, these experiences occur simultaneously but they are not co-conscious as they occur in separate centres of consciousness. The question for Hurley is; how do we account for the difference between this case and a case where two states or experiences are co-conscious—how do we give a constitutive account of the co-consciousness at a time of two or more mental states?

Hurley's (1998b) approach consists of examinations of some empirical results achieved in experiments conducted with so-called split-brain patients (that we encountered briefly in chapter one) and also considers some hypothetical cases. 'Split-brain' patients fall into two categories: a commissurotomy patient is one who has the bundle of fibres that connect

the two hemispheres of the cortical brain—called the corpus callosum—severed (which is a recognised treatment for epileptic seizures); occasionally the corpus callosum is congenitally absent, this is referred to as *callosal agenesis* and those subject to this condition are sometimes referred to as *acallosals*. Under experimental conditions, when a commissurotomy patient or an acallosal fixes his gaze, information from the left visual field reaches the right hemisphere of the brain only and information from the right visual field reaches the left hemisphere of the brain only.

For example, the savant Kim Peek, on whom the film *Rain* Man is based, is an acallosal and can read two pages of a book simultaneously—one with each eye—doing so with near perfect information retention. So, given the absence of the corpus callosum in a patient, an experimenter can send differing information to each hemisphere by displaying different stimuli in the left and right visual fields. Thus, under laboratory conditions, the experimenter can induce inconsistencies of content, or disunity, in the commissurotomy patient or acallosal.

For example, suppose the experimenter simultaneously projects a single point of red light to the patient's right visual field and a single point of green light to the patient's left visual field. Suppose also that when asked what he sees while fixating the patient responds using the right hand (controlled by the left hemisphere) that he sees *just one point* of light that is wholly red and *simultaneously* responds using the left hand (controlled by the right hemisphere) that he sees *just one point* of light that is wholly green. In such a case the experimenter will have induced a normative incoherence in the mental content of the patient—it does not make sense for there to be an experience of just one point of light, at one time, that is both completely red and completely green.

Normative constraints such as these are redundant when we consider the separateness of Otto's consciousness from Inga's consciousness at a time because the separateness of their consciousnesses already seems to be over-determined by other factors like their bodily separateness. If Otto is thinking *that p* at t and Inga is thinking *that not-p* at t, then we do not need to appeal to a constraint of normative coherence to distinguish the two separate centres of consciousness because we can simply appeal to their bodily separateness. So constraints of normative coherence are redundant when considering cases of disunity between two separate bodies.

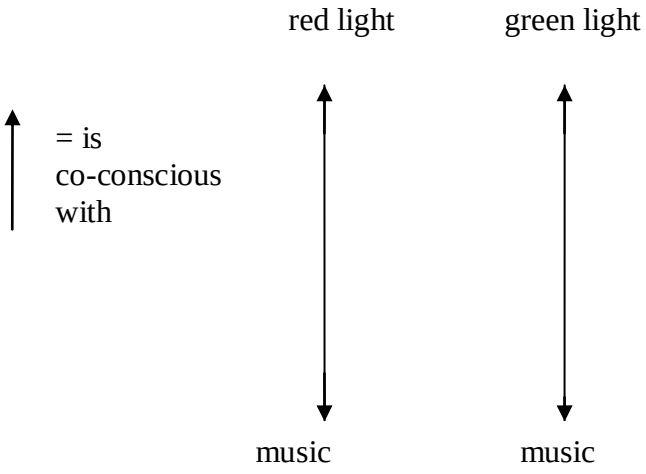
Norms do have a role to play, however, in cases of possible disunity harboured within one body. This kind of situation can arise in cases of multiple personality and in cases of commissurotomy. Consider two mental states *a* and *b*: mental state *a* has the content *that p* and mental state *b* has the content *that q*. Furthermore, *that q* entails *that not-p*. It is impossible then (in general) for mental state *a* and mental state *b* to be co-conscious because a person cannot be thinking both *that p* and *that not-p* at the same time. Likewise, if a person responds, under experimental conditions, using the right hand, that he sees just one point of light that is wholly red and simultaneously responds using the left hand that he sees just one point of light that is wholly green, then we can deduce that consciousness is not unified in this case. We cannot, given constraints of normative coherence, attribute two such inconsistent contents to one consciousness, so we segregate the contents or recognise disunity. That is, the patient in the hypothetical experiment just considered has manifested disunity of consciousness.

Hurley emphasises she is not claiming that this normative constraint, which forbids inconsistent or incoherent contents in one consciousness at a time, applies universally or indefeasibly. She admits that there may be

cases where it *does* make sense to attribute inconsistent contents to one consciousness at a time. Her point is simply that a constraint of normative coherence and consistency does apply in many cases. And she states that the role of such a constraint is demonstrated by the way in which disunity is attributed in neuropsychological cases such as the commissurotomy case considered above.

So norms of coherence can account for separateness in cases such as the commissurotomy one considered above but they can also help us in making sense of duplicate states of consciousness. Suppose that while our hypothetical experiment is being conducted there is music playing. In this case both left and right hemispheres will be conscious of the music. Nevertheless, it will still remain the case that the right hemisphere is conscious of a green light and not a red light and the left hemisphere is conscious of a red light and not a green light. It is difficult to make sense of such a situation. One way explored by Hurley is to posit two separate centres of consciousness in which conscious content, in this case consciousness of the music, can be duplicated.

Duplication of content in separate centres of consciousness



Here, consciousness of the music is duplicated in each centre of consciousness—each hemisphere. When the experiment ends and the inconsistency that had been induced also ends, the duplication of conscious contents in each hemisphere may continue. However, norms can only account for the duplication of contents in this way in cases of normative inconsistency. Norms cannot account for separateness where there are no normative inconsistencies or incoherence between contents. Consequently, normative constraints cannot account for the possibility of the global duplication of contents over time. To put it another way, norms cannot account for the wholesale duplication of contents of one consciousness simultaneously in another consciousness. Thus, a constitutive account of the unity of consciousness at a time requires, according to Hurley, moving beyond the rules of normative coherence.

The Vehicle/Content Distinction

So far we have seen that although normative constraints do play a role in a constitutive account of the unity of consciousness for Hurley, normative coherence is not sufficient for unity as there can be normative coherence between contents realised in two distinct consciousnesses and, furthermore, normative coherence cannot account for the possibility of the global duplication of contents.[23] We need, then, a complement to these normative considerations. The obvious place to look for this, says Hurley, is at the subpersonal level of the vehicles of content.

Hurley says that the internalist position that vehicle externalists claim to challenge holds that the vehicles of content are exclusively brain

[23] There are more problems with attempting to rely on normative coherence to offer a constitutive account of unity: an account based purely on normative coherence lacks the resources to distinguish cases of partial unity from cases of duplication and also trying to account for unity in this way can give norms too much power over time as we may need to project isolated incidences of normative incoherence through time and postulate two separate centres of consciousness in an otherwise normal subject.

structures and processes. If this internalist position were true then Hurley claims that one would expect the neurophysiological structure of the brain to be reflected in the personal level structure of consciousness. That is, disunity of brain structure should manifest itself in disunity at the personal level, the level of consciousness. This would explain why commissurotomy patients, such as the patient considered in our hypothetical experiment, manifest disunity of consciousness; because there is disunity of brain structure given that the connection between the two hemispheres of the brain has been severed. But relying on neuroanatomical isomorphism in this way to account for the unity of consciousness at a time has problems in certain cases, according to Hurley.

In ordinary cases the corpus callosum is present and intact and consciousness is unified, suggesting, possibly, that neurophysiological isomorphism may be a good candidate for the necessary subpersonal element for a complete account of the unity of consciousness. However, the corpus callosum is also present and intact in multiple personality patients yet consciousness is not unified. In multiple personality cases there seem to be separate centres of consciousness and there may be duplicate conscious states at a given time. Thus, Hurley says that neuroanatomical structure does not reflect the structure of consciousness here. In cases of commissurotomy, where the corpus callosum has been severed, responses that manifest disunity can be elicited from the patient, as in our hypothetical experiment. However, displays of disunity in the commissurotomy patient are extremely rare outside of artificial, experimental conditions such as these. When the commissurotomised subject is observed under normal conditions he/she is usually perceptually and socially competent and indistinguishable from ordinary cases. So, in cases of commissurotomy we have a clear anatomical, subpersonal basis for attributing disunity at the personal level but this

disunity, in the form of normative incoherence at the personal level, can be very difficult to induce and rarely manifests itself outside of experimental conditions.

Furthermore, the inability to integrate visual information, for example, between one hemisphere and another is "biologically trivial" (Kinsbourne 1974: 253). This is because slight changes in orientation can enable the patient to distribute the same information to both hemispheres. Such changes in orientation are called access movements and are reliable methods of information integration even where alternative methods such as cross-cuing fail. Cross-cuing is deliberate communication between hemispheres by means of an alternative, usually external, pathway. A smile can function as a method of cross cuing as it crosses the vertical axis of the face and thus relays proprioceptive information to both hemispheres.

In cases of callosal agenesis the corpus callosum is congenitally absent, yet it is extremely difficult, even under experimental conditions to elicit responses that are normatively incoherent and, thus, indicative of disunity. Despite the fact that callosal agenesis patients are anatomically similar to commissurotomy patients they pass almost all of the tests that commissurotomy patients fail. This illustrates, says Hurley, that unity of anatomical structure is not necessary for unity of consciousness. One could argue that there are other neurological structures, such as subcortical and ipsilateral structures, that could transfer information between the hemispheres. But these structures are also present in commissurotomy patients—the presence of these structures is not necessarily what makes the difference says Hurley.

It does not seem correct to attribute disunity to a callosal agenesis patient who does not manifest disunity under testing. If the patient displays unity

under testing then it is surely a mistake to attribute disunity on the basis of the absent corpus callosum alone (Hurley 1998b). Unity of consciousness is a personal level phenomenon and the anatomical distinction that we can draw between unified and disunified brain structures is actually independent of the intuitive distinctions that we draw between normal cases and non-normal cases such as multiple personality, commissurotomy and callosal agenesis. The intuitive distinctions that we draw between unified and disunified consciousnesses are based on norms applied to actions. If these intuitive distinctions are independent of brain structure then the actions to which the norms are applied, and on the basis of which we make our intuitive distinctions, cannot be regarded as necessarily providing evidence for underlying brain structure. Unity at the subpersonal level in anatomical brain structure can yield disunity at the personal level, whereas unity in consciousness at the personal level can manifest despite disunity at the subpersonal level in anatomical brain structure. The lesson to be learned here, according to Hurley, is that it is a mistake to assume that properties of the subpersonal level of neuroanatomical structure project onto the personal level of contents and consciousness, or vice versa.

This is a warning to heed the vehicle/content distinction as also noted by Dennett (1991), Dennett & Kinsbourne (1992), Millikan (1991) and others. For example, nobody thinks that a vehicle that enables particular phenomenal content of the type ‘green’, or ‘green there!’ must itself be green, nor should anyone claim that the syntactic, recombinant (classical) structure of thought must be implemented by a classical architecture. In computational terms, it is completely uncontroversial that 0s and 1s can represent information other than 0s and 1s. In short, it is a mistake to assume that the properties of vehicles project onto contents, and vice versa.

The above considerations raise the question of how commissurotomy patients and acallosals integrate information so as to pass experimental tests of unity, despite disunity of neuroanatomical structure. Furthermore, how do acallosals integrate information so as to pass the experimental tests of unity that commissurotomy patients fail? One way may be reliance on ipsilateral and subcortical neural pathways. This is more than likely to be the case where acallosals pass tests of unity that rule out the use of external mechanisms of integration such as access movements or cross-cuing. But if, instead, acallosals employ external mechanisms of integration are we justified in attributing disunity? If we are tempted to attribute disunity to the acallosal in this case but not in the previous case then we must specify the relevant difference. We cannot simply point to the fact that unity is achieved in the former case by brain structure because, as we have seen, unity of neurophysiological structure is often not what matters in non-normal cases. So, Hurley offers a hypothetical case for consideration. In a move similar to that made by Clark and Chalmers in presenting the Parity Principle, she considers the significance of the distinction between internal and external methods of integration; firstly for recently operated commissurotomy patients, and secondly for callosal agenesis patients.

If a recently operated commissurotomy patient uses an external means of integration when access movements are prevented then this may well provide evidence for a division in consciousness. But similar dependence on cross-cuing by an acallosal will not necessarily provide evidence for a division in consciousness. In the case of the acallosal this behaviour will be necessary throughout the development of his conscious mental capacities. For him there is no structure of consciousness, unified or disunified, prior to the use of external mechanisms of integration. The

structure of consciousness, in this case, is something that emerges along with the external mechanisms of integration. The external mechanisms of integration, for this acallosal patient, are temporally and ontologically prior to the unity of consciousness that eventually manifests itself at the personal level.

Consider, next, two hypothetical acallosal patients. The first displays a high degree of integration and coordination between hemispheres that is achieved in everyday circumstances by *external* mechanisms. The external mechanisms employed here *would* suggest communication between two separate centres of consciousness along with duplication of some contents *if* the patient were a recently operated commissurotomy patient (Hurley 1998b). The second patient displays a high degree of integration and coordination between hemispheres achieved in everyday circumstances by *internal* mechanisms such as subcortical and ipsilateral pathways.

Hurley claims these acallosal cases display parity; in neither case are the mechanisms of integration a means of communication between two separate centres of consciousness with resulting duplication of contents. This is because in the cases of the acallosals there was no structure of consciousness prior to the mechanisms (internal or external) of integration. Similarly, it is not correct to say that the corpus callosum is a means of communication between two separate centres of consciousness with resulting duplication of contents in a normal case. Thus, the same claim for parity cannot be made for adult, recently-operated commissurotomy patients.

> For acallosals, to the extent either internal or external mechanisms of integration function reliably; there is no reason not to regard them as

> part of the vehicles of co-conscious contents and of a unified consciousness. (Ibid: 191)

To explain this in more detail Hurley offers an analogy with perceptual adaptation to up-down-inverting lenses (see Hurley 1998b: 191). She claims that acallosals stand to commissurotomy patients with respect to the unity of consciousness as persons who have always worn up-down inverting lenses stand to adults who put on inverting spectacles in perceptual adaptation experiments with respect to type of conscious experience. Persons wearing up-down inverting spectacles are found, after a period of time that involves dynamic action in the world, to adapt to the wearing of these spectacles such that the world 'rights itself' again and normal 'right side up' perceptual experience is restored. But for the subject who wears these inverting lenses from birth the situation is slightly different. Although similar means of up-down re-routing may be used to achieve reliable perception of the environment, these methods, for the latter subject, are a pre-condition of correct perception of the world. Now, whether the up-down re-routing is achieved via internal or external mechanisms in this case is irrelevant according to Hurley. What is relevant is that up-down rerouting is achieved and reliable perceptual experience is accomplished. Thus, just as it makes no difference here how the up-down rerouting is achieved, it makes no difference in principle whether the mechanisms of integration for the callosal agenesis patient are internal or external says Hurley.

What does matter in cases such as these, says Hurley, is that the mechanisms of integration 'go subpersonal'. That is, what is essential for achieving unity of consciousness by means of these methods is that the mechanisms (be they internal or external) should not normally be subject to control at the personal level. They can initially be subject to control at the personal level for the commissurotomy patient who has recently had

his corpus callosum severed. But unity will not be achieved until control of these mechanisms goes subpersonal such that unity is achieved automatically and information is integrated between the two hemispheres without conscious effort. This condition echoes that offered by Clark and Chalmers in relation to cases of genuine external memory: Otto's case was not to count as a genuine case of belief unless his access to the notebook was transparent and automatic; or subpersonal.[24]

Hurley's empirical and hypothetical cases demonstrate that commissurotomy disrupts an established subpersonal basis for the structure of consciousness and it is reasonable to suppose that new external links would at first be treated as foreign, rather than as simply re-establishing prior unity. But the commissurotomy patient could adapt over time such that external mechanisms of integration acquire the status of a subpersonal basis for the unity of consciousness rather than a means of communication between separate consciousnesses. Thus, says Hurley, there is no reason in principle why external mechanisms of information integration could not contribute to supporting a unified consciousness and, in some cases, external mechanisms may be necessary to enable the unity of consciousness.

The difference between internal and external mechanisms of integration is not the same as the difference between an automatic, subpersonal process that underwrites one unified consciousness, and a conscious personal-level coordination between two separate consciousnesses. Subpersonal processes that underwrite unity are not necessarily internal on Hurley's account. There is nothing wrong with the idea that an external mechanism of integration could be part of the system that supports the unity of consciousness. This idea, says Hurley, appeals to a

[24] But as we noted in chapter one, this didn't seem a plausible condition to place on Otto's access to his notebook.

system of causes and effects in the same way that an explanation citing purely internal mechanisms of integration would, it's just that some of the causal routes go external.

> Whether the causal mechanisms are wholly internal or partly external may be incidental to the unity of consciousness: context and history of development may matter more. (Ibid: 193)

So, Hurley concludes that there need not be neuroanatomical unity underwriting consciousness. What seems to be critical for unity is the functional structure and history of a dynamic causal system, which may have external as well as internal elements. Thus, for Hurley, the neuroanatomical isomorphism suggestion fails; the unity of consciousness at a time does not necessarily reflect an isomorphic unified neuroanatomical structure and the structure of consciousness, like the contents of consciousness, is as Hurley puts it: "easy going; it will take a ride wherever it finds one" (Ibid.: 194). If so, then an appeal to neuroanatomical isomorphism will not provide the subpersonal component of an objective account of unity that is needed. But Hurley suggests that the notion of a dynamic singularity may be able to succeed where neuroanatomical structure fails.

A dynamic singularity is a structural singularity in the field of causal flows realised at the subpersonal level of explanation. This singularity is characterised through time by a tangle of multiple feedback loops, each with varying and distinctive orbits. It is centred on the organism and moves through the environment with the organism. But although the dynamic singularity is centred on the organism it can extend beyond the organism because the causal feedback loops extend beyond the boundaries of the organism and into the environment just as in the case of the acallosal who uses external mechanisms of integration. The dynamic

singularity, unlike the organism, does not itself have any clear, unchanging boundaries.

Now, although the realisation of the causal roles that form the structure of the dynamic singularity would involve neurophysiological states: "it would not consist in mere neuroanatomical isomorphism: the structure of the realizing neurophysiological states need not correspond to the structure of consciousness" (Ibid: 207). This, in certain cases, may be precisely because some of the causal loops of the dynamic singularity go external. For example, the hypothetical acallosal patient discussed earlier has used external methods of integration since birth and depends on these methods for the unified structure of his consciousness. For him, there was no structure of consciousness prior to these external mechanisms of integration; thus, there is no reason to deny that these external mechanisms are part of the vehicles of co-consciousness.

So, Hurley suggests that normative coherence and dynamic singularity may be individually necessary and jointly sufficient for the unity of consciousness at a time. Moreover, the dynamic singularity can extend beyond the boundaries of the organism. If this is the case then the subpersonal vehicles that realise a unified consciousness can extend into the environment.

This does not mean, however, that *consciousness* extends into the environment. Rather, the claim is that these extended vehicles provide the structural basis that enables the unity of consciousness at a time. And, given that we should be wary, in accordance with the vehicle/content distinction, of projecting properties from the vehicles of content onto the content itself it would be poor judgment to simply assume that the locational properties of the vehicles of consciousness project onto consciousness itself.

Hurley's argument for the extension of vehicles into the world does not stop at the necessity of the dynamic singularity in accounting for the unity of consciousness. She moves beyond these considerations to explore one of the traditional tools of the externalist; Twin Earth thought experiments, in order to present further support for her Vehicle Externalism.

In chapter one I briefly mentioned an argument presented by Clark and Chalmers (1998) that is introduced in order to lend weight to their version of EMT and emphasise its distinctness from content externalism. The argument went as follows: content externalists frequently rely on Twin Earth thought experiments in order to demonstrate that when internal factors are held constant mental content type can still change along with environmental changes. However, in cases such as these (like the water/retaw case) the behaviour of the doppelgängers is identical, indicating that content externalism is a passive form of externalism that has no consequences for action. Active Externalism, however, employs external factors that have an ineliminable role to play in the production of behaviour in the here and now as in the case of Otto and Twin Otto:

> [W]e can even construct the case of Twin Otto, who is just like Otto except that a while ago he mistakenly wrote in his notebook that the Museum of Modern Art was on 51st Street. Today, Twin Otto is a physical duplicate of Otto from the skin in, but his notebook differs. Consequently, Twin Otto is best characterized as believing that the museum is on 51st Street, where Otto believes it is on 53rd. In these cases, a belief is simply not in the head. (Ibid: 14)

This seems initially like a straightforward extension of the scope of the Twin Earth thought experiment to argue for the extension of factors that

can affect behaviour in the here and now rather than just passive and distal factors that, while affecting content type, make no difference to the production of behaviour. But Clark and Chalmers' argument here seems to rely on the thought that Twin Otto can mistakenly write '51st street' in his notebook and yet remain a physical duplicate of Otto from the skin in. That is, if internal factors are held constant while external factors are altered, as in the Twin Otto case, so as to yield different behaviour then in some cases external factors will have an ineliminable role to play in the production of behaviour and mental states may extend into the world.

It is not clear though, how Clark and Chalmers propose that Twin Otto is to remain a physical duplicate of Otto from the skin in given that he writes '51st street' in his notebook, rather than '53rd street'. Presumably, Twin Otto's writing a different number will require different hand movements than Otto's, which will require a different pattern of neural firings, and when consulting his notebook the pattern of light striking Twin Otto's retina will be different to the pattern of light striking Otto's retina, thus the signals sent to the brain will differ and the physical internal states will differ.

It may well be the case that what is written in the notebook plays an ineliminable role in the production of behaviour in each case but since it seems that Otto and Twin Otto are not in fact physically identical from the skin in, the door is open for those who would argue that the different beliefs in each case is *constituted* solely by internal factors although it may be partially *explained* by external factors. I will return to the Twin Otto case periodically in the course of this section as I examine in detail Hurley's arguments relating to Twin Earth thought experiments of this sort.

Clark and Chalmers' mistake here is an example of what Hurley (1998b) calls the general duplication assumption, a view that she says is typically associated (though not exclusively) with internalism. On the internalist view, our *internal* physical brain states fix the contents of our beliefs and experiences, and so mental content depends only instrumentally on the external environment. That is, mental content depends on the environment only to the extent that it can affect or change the internal physical states that fix mental content. Thus, a person's mental content, on the internalist view, is, in an important sense, autonomous or independent of the environment.

So, if mental content is internal and autonomous of the external world and if changes in the world affect mental content only instrumentally by affecting changes in internal physical states and processes, then the duplication of mental content by means of the duplication of internal physical states and processes should be possible in radically different environments. This internalist position regarding the duplicability of mental contents has, as we have seen, received widespread challenge in the form of content externalism. However, as Hurley points out, the duplicability of internal *physical* states and processes in radically altered environments is often taken to be unproblematic. To believe in the possibility of general duplicability is to believe that, in general, internal physical states can be duplicated in radically different environments.

Hurley argues that this assumption of general duplicability is widespread; it is the default starting point for those wishing to theorise about mind, although it is not widely recognised and as a consequence it is not widely challenged. This is evident in consideration of Clark and Chalmers' Twin Otto thought experiment. As we have seen, Clark and Chalmers' argument only works if Otto and Twin Otto are physically identical from the skin in; that is, if Otto's internal physical states can be duplicated in

an altered environment—one where he mistakenly writes '51' instead of '53'. But it is difficult to see how identity of internal physical states can be maintained between the doppelgangers given the altered nature of Twin-Earth in this case. Nonetheless, Clark and Chalmers take the duplication of internal physical states to be unproblematic here. Indeed, the success of twin-earth thought experiments depends on the possibility of duplicating internal physical states in counterfactual worlds. Content Externalists who wish to demonstrate that mental content can alter although internal physical brain states remain constant often support their arguments by constructing Twin Earth thought experiments, which are predicated on the assumption that the duplication of internal physical states in altered environments is not problematic.

Hurley argues that assuming that duplication of this sort is in general unproblematic is a mistake. She states that challenging the general duplication assumption is a matter of scrutinising the presupposition of duplicability in specific cases. And this challenge to the general duplication assumption leads, ultimately, to the suggestion of externalism about the vehicles of content.

Now, although adherence to the assumption of general duplicability is a mistake, there are certain cases for which duplicability would seem to hold without a problem. Consider the Twin Earth thought experiment where Twin Earth has a substance called 'retaw', which has a different chemical structure to water. Here, duplicability seems possible without too much difficulty. That is, it may be possible to have a Twin Earthling on Twin Earth that is a physical duplicate of the Earthling on Earth from the skin in.[25] This duplication of internal physical states would seem to

[25] We can ignore here, for the purposes of argument, the fact that the human body would be made up of 70% water whereas the Twin-Earth body would be made up of 70% retaw, ruling physical duplication out in this case.

hold despite the radical change to the environment where water is replaced with 'retaw'. The point here, however, in challenging the general duplication assumption is not to try and show that duplication is always and everywhere impossible. Rather, the point is to challenge the *general* duplication assumption, to challenge the idea that it is always, in every case, possible to radically alter the environment while duplicating internal physical states. To challenge the general duplication assumption means to challenge the idea that it makes good sense in principle to suppose that internal physical states can be duplicated in radically different environments—whether duplication holds in particular cases should be an empirical matter.

Certainly, duplication may not be problematic in certain uncomplicated cases; however, as we saw in the Twin Otto case, as the examples increase in complexity and involve action in the world by an agent, duplication becomes increasingly difficult. In these problem cases we need to consider not just internal operations over inputs that lead to outputs but we also need to examine certain functions from output back to input. Certain of these functions operate internally such as efferent feedback and proprioceptive feedback, whereas others operate externally, via the environment of the agent, like visual feedback from movement. Hurley warns against a tendency that she observes in philosophy to simplify thought experiments so as to gloss over such things as feedback relations. She says that in doing this we are in danger of making 'toy' agents, who are not true agents at all. We need to consider, instead, real agents acting dynamically in their environments. Doing so can make problems for duplication.

Thus, once we consider Otto and Twin Otto as dynamic agents acting in the world and not as 'toy' examples frozen at a point in time, we see that the duplication of their internal physical states becomes at the very least

problematic, if not impossible. One option in trying to preserve the possibility of duplication is to introduce 'reversing devices' that can negate the changes brought about to internal physical states by the radically altered environment, thus maintaining the duplicability of internal physical states. We will examine possibilities such as this, along with Hurley's response, in the next two sections.

Colour Inversion and Left-Right Reversal

Hurley begins her argument against the general duplication assumption (which leads, she argues, to externalism about the vehicles of content) by examining a red/green inversion case in which duplication seems to be unproblematic. Everything that is green on Earth is red on Inverted Earth and vice versa. Furthermore, the people on inverted Earth call red things 'green' and green things 'red'. Other than this difference, Earth and Inverted Earth are twin possible worlds. Now, suppose we have an Earthling and a Twin Earthling, who both interact with their worlds in the same manner: will their internal physical states be duplicated? It seems not, for when the Earthling encounters a green object on Earth his Twin Earthling encounters a red object on Inverted Earth. Thus, the light striking the retina of the Twin Earthling will be different to the light striking the retina of the Earthling, the signals sent to the brain will differ, and consequently the physical internal states will differ.

Hurley states that: "This is where the causal slack between outside world and internal physical states is exploited" (1998b: 299). We suppose that, in principle, a change can be made to the causal pathways that connect the outside world to the Twin Earthling's internal physical states. This change would vindicate the assumption of duplicability in this specific case by cancelling out or compensating for the differences that are on Inverted Earth as compared with Earth. In this case of a red-green

Inverted Earth we may suppose that colour inverting goggles are fitted to our Twin Earthling. These goggles, obviously, fall outside of the boundary that marks off the internal physical states and processes that are to be duplicated from the 'outside' world. If we used colour inverting lenses inserted into the retinas of the Twin Earthling instead then it could be objected that the lenses are within the boundary of the skin and would themselves have to count as part of the internal physical state that is supposed to be duplicated between the two possible worlds. As with Clark and Chalmers' thought experiment, the skin seems to be the generally accepted boundary in cases like this, but as we will see it may be necessary to retreat further inwards to the boundary around the central nervous system (CNS), for example, if duplication of internal physical states is to be preserved. Hurley points out that: "[t]his looks like a harmless enough piece of boundary-drawing, but the gerrymandering that begins here ultimately turns problematic" (Ibid: 300).

With the colour inverting goggles fitted to our Twin Earthling the causal pattern that holds on Earth between certain internal physical states and green objects in the environment now holds on Inverted Earth for red objects and vice versa. That is, perception of a red object on Inverted Earth by our Twin Earthling will cause the same internal physical state as when our Earthling perceives a green object on Earth. In this case, and other simplistic cases, the achievement of duplicability seems to be unproblematic, at most it raises technical questions such as how the colour inverting lenses would work—it does not raise difficulties in principle with the possibility of duplication.[26]

[26] However, Hurley points out that the fact that the inversion in this case is symmetrical is crucial. Non-symmetrical inversions would be subject to what she calls the search-and-replace-fallacy and, consequently, duplication may not be possible for non-symmetrical inversions. Thus, there is a problem for the general duplication assumption even before we consider issues relating to action and dynamic feedback. More on this point later.

But Hurley claims that switching from ‘what’-system perception cases—involving colour, details or textures—to considering ‘where’-system perception cases—involving where things are and how fast they are moving—makes duplication more difficult. In particular, the action and dynamic-feedback that is intimately bound up with ‘where’-system perception makes duplication difficult. She is not claiming that duplication is always unproblematic for ‘what’-system perception and never for ‘where’-system perception but does claim that there is something about ‘where’-system perception that makes duplication difficult—‘where’-system perception involves orientation and localization that involves acting in the world more intimately than colour-perception does.

Hurley asks what would happen if we modelled a left-right reversal case on the red/green inversion case we considered earlier but this time taking action and dynamic feedback into account. This would introduce real-world complexities into the example, protecting it from the over-simplicity charge which could be levelled at other twin earth thought experiments. The problem posed by ‘where’-system perception for the possibility of duplication in cases such as this becomes evident in the requirement for a reversing device that cancels the left/right inversion. Such a device, as in the colour inversion case, must yield the same neural input to the Twin Earthling as occurs for the Earthling, and must be placed outside of the boundary within which duplication is to be preserved. If the device must be placed inside this boundary in order to achieve identity of neural input then central states will differ between the Earthling and Twin Earthling and duplication will not be achieved.

Experiments carried out by Taylor (1962) with the mathematician Seymour Papert concerning the perceptual adaptation of Papert to the long-term wearing of left-right reversing goggles provide interesting

support for Hurley's thought experiment here. Left-right reversing goggles cause objects that would normally appear on the right of the subject to appear on the left, and vice versa. Papert underwent a program of training while wearing the goggles each morning. In the evening the goggles would be removed and he would go about his daily activities as normal. During one of the training programs Papert experienced, for the first time while wearing the left-right reversing goggles, the perception of an object in its true position:

> But it was a very strange experience, in that he perceived the chair as being both on the side where it was in contact with his body and on the opposite side. And by this he meant not just that he knew that the chair he saw on his left was actually on his right. He had that knowledge from the beginning of the experiment. The experience was more like the simultaneous perception of an object and its mirror image, although in this case the chair on the right was rather ghost-like. (Taylor 1962: 202)

Bear in mind that Papert only wore the goggles during the morning and returned to normal vision in the evening. Despite this, he adapted to the wearing of the goggles quite quickly and the adaptation was total, that is, he was, following training, able to perceive objects as normal, as if he was not wearing the reversing goggles. So Papert's adaptation to the wearing of the goggles was not impeded by the fact that he returned to normal perception each afternoon. In fact, Taylor reports that Papert's adaptation to the wearing of the goggles had been faster than in two previous experiments that had been conducted. This was due, according to Taylor, to the systematic training that Papert underwent in the third experiment. Furthermore, following adaptation, Papert was able to put on and remove the goggles without changing the left-right ordering of his perceptual field and without his behaviour being disrupted:

> This was strikingly illustrated when the subject rode a bicycle while wearing the spectacles, and took them off and replaced them without changing course or wobbling or showing any other signs of disruption. Objects that he perceived as being on his left while wearing the spectacles were still on his left when he took them off. (Ibid: 204)

In Hurley's left/right reversal thought experiment we are given Seymour, who is on Earth, and enjoys riding his bike while wearing his (normal—no reversing lenses included) biking goggles. While on Twin Earth Twin Seymour also enjoys riding his bike but in a world that is the mirror image of Seymour's—a global left-right reversal has taken place. The question is: is it possible to duplicate Seymour's inner physical states on Earth in Twin Seymour on Mirror Earth and, if so, how? If the general duplication assumption is correct then duplicating Seymour's internal physical states in Twin Seymour should not be a problem.

As things stand if Seymour sees a sphere on his left then Twin Seymour will be seeing a sphere on his right, thus the light entering the retinas of Seymour and Twin-Seymour will be different and the neural signals being sent to the brain of each will differ. It follows, then, that the internal physical states differ and duplication is not achieved. In order to get duplication we need some cancelling or compensating reversing devices as we did in the red/green inversion case. Hurley suggests left/right reversing goggles such as Papert wore in the real experiments considered above. These goggles must, of course, be outside the boundary within which we are trying to achieve duplication. Otherwise duplication will fail as Twin Seymour's goggles must be physically different from Seymour's in order to perform the necessary reversal.

Hurley says that wearing these goggles could possibly achieve the desired duplication—but only on condition that both Seymour and Twin

Seymour stand still. When Seymour sees a sphere on his right, Twin Seymour will now also see a sphere on his right. So the neural signals to each from the eye to the brain would be the same and duplication would seem to be preserved. But note again here the tendency towards oversimplification in thought experiments. In order to avoid this objection we must examine the individuals as functional individuals within their environments: "[A] stationary perceptual system is a truncated and denatured one; it's not significant here that you can trick it" (Hurley 1998b: 304).

Once Seymour and Twin Seymour start to move then duplication is lost. Imagine that Twin Seymour has his left-right reversing goggles on and both Seymours are seeing a sphere on their right. Duplication will be lost as soon as Twin Seymour tries to reach out and point at it, for example. True, it will appear to him with the reversing goggles on that he is reaching toward the sphere with his right hand say, but in reality he will be reaching with his left hand and proprioceptive signals will inform him of the discrepancy. Like Papert in the real experiment Twin Seymour would, doubtless, learn to adapt and would be able, like Seymour (and Papert) to remove and replace his goggles while cycling and continue to function normally. But this isn't the issue here, what we want to know is whether it is possible to duplicate Seymour's internal physical states on Earth in Twin Seymour in a mirror reversed environment. And, given proprioceptive feedback and the fact that we need Seymour to be able to engage with and function in the world we are going to need more reversing devices than the left-right reversing goggles here.

So, in order to achieve duplication here we would need some sort of motor reversal device that would cause Twin Seymour's right hand to move rightward when Seymour's left hand moves leftward, while at the same time preserving the internal physical duplication. These differences

in which hand is used and in which direction it reaches must be outside the boundary within which physical states are to be identical, just as the goggles are, otherwise our duplication attempt is already defeated. But likewise, the device that causes these changes must also be outside the boundary. Thus, we are faced with a bit of a dilemma: if we don't switch from one hand to the other (when Twin Seymour is pointing at the sphere say) then we won't duplicate visual feedback from hand movements but if we do switch hands then we won't duplicate muscle contractions and motor nerve firings (as Twin Seymour will be moving the opposite arm to Seymour) but this means, says Hurley, that the boundary within which we wish achieve duplication cannot be the skin – if we wish to preserve duplication we will have to retreat further inwards to the boundary of the CNS.

To summarise, if we are to achieve duplication in this case we are going to need at least two things besides the reversing goggles, we will need a motor reversal device and we will need a proprioceptive reversal device. It is difficult to see, in either case, how we could fulfil the need for these reversal devices without altering physical states *within* the body. This is particularly the case for the proprioceptive reversal device. However, if we do have to alter physical states within the body in order to fulfil the need for the reversing devices then we are changing the internal physical states that are supposed to be duplicated, the states that are supposed to be held constant cannot, then, be held constant. In this case duplication does not seem to be possible. However, if we retreat from the skin as the boundary within which we wish to duplicate physical states to the CNS then both the motor reversal and proprioceptive reversal devices can be moved inside the skin to the boundaries of the CNS, changing the signals on the way out. Thus, Twin Seymour's motor reversal device would take the motor signals that emerge from the CNS and switch them to the opposite side—reversing the brain's contralateral control. The result of

this is that the right hand moves rightward instead of the left hand moving leftward. Similarly, the proprioceptive reversal device now takes signals coming from the arms, for example, and switches them to the opposite side. The result is that the proprioceptive signals coming from the right arm now input to the right hemisphere of the brain instead of the left and vice versa. Both of these devices, although within the skin, must be located outside of the CNS if duplication within the CNS is to be achieved.

Even with these reversal devices implanted, however, there is a problem when Seymour removes his goggles on earth. If Twin-Seymour does likewise on Twin-Earth then duplication will not be preserved as the reversing effect of the goggles is removed. This is easily overcome, however, by simply moving the goggles inward—surgically inserting left-right reversing lenses in Twin Seymour's retinas. So, Hurley concludes that in this case duplication is possible provided we move the boundary inward from to the skin of the subject to the CNS of the subject and keep all reversal devices within a *safe-zone* free from the possibility of tampering by the agent. But this safe-zone strategy is not likely to work in all cases.

> Can we find a case in which the safe zone strategy will not work? Programmatically, what would be needed would be a situation in which something analogous to putting the reversing goggles off and on is a normal activity and so cannot be avoided by the safe zone strategy. (Ibid: 314)

So, the tactic employed by Hurley here is to illustrate how the mirror reversal case forces us to retreat to the CNS as the boundary within which physical states are to be duplicated in order to preserve the possibility of duplication. It also suggests to us that the retreat to the CNS in order to

preserve the possibility of duplication may not always work. Once we introduce real-world complexities into our inverted earth thought experiments we may be forced to alter the very internal states that we are trying to duplicate. To pursue this line of enquiry we will examine cases which challenge the general duplication assumption further.

El Greco Worlds

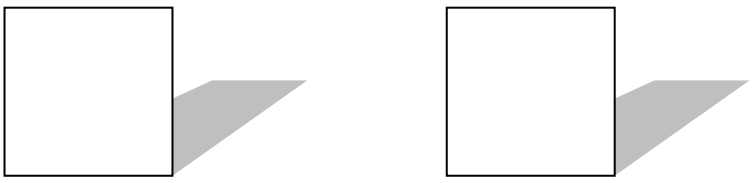
In discussing the red/green inversion case we focused on 'what'-system perception and in the mirror reversal example the focus was on location and orientation; 'where'-system perception. Shape is a 'what'-system property but shape distortion can have consequences for both 'what' and 'where' system perception. This sort of shape distortion occurs on what Hurley calls El Greco worlds. Worlds that are exactly similar to earth, El Greco worlds are twin earths except for one property; a vertical stretch has been applied to everything in the direction opposite to the pull of gravity.

Hurley's subjects for her El Greco thought-experiments are Dom and Twin Dom and our duplication boundary is around the CNS, not around the whole person. So in this case, because your body is one of the things that you perceive, Twin Dom's body will also have the vertical stretch applied to it, although this stretch will not apply to his CNS. Hurley's strategy in arguing against general duplication is to consider different kinds of activity for Dom on Earth and Twin Dom on Twin Earth, where the one-off vertical stretch has been applied. Consideration of certain kinds of activities on Earth and Twin Earth will provide cases where the assumption of duplicability seems untenable and ultimately leads to suggestions for the extension of vehicles beyond the boundaries of the CNS and the skin and into the world.

Suppose that at t_0 Dom is physically identical to Twin Dom and Earth is physically identical to El Greco Earth. Now, suppose that at t_1 a one-off stretch is applied to all physical objects on El Greco Earth, including bodies, therefore, the stretch will apply to Twin Dom's body. Also, the stretch is proportionate; such that at t_1 all objects on El Greco Earth, including Twin Dom's body, are twice as tall as their counterparts on Earth. But the stretch does not apply to the CNS, so that Twin Dom's CNS remains a physical duplicate of Dom's CNS at t_1.

On this El Greco world the stretch does not entail a simple doubling of the size of objects. The stretch applies only in the opposite direction to gravity. So, the El Greco Earth counterpart of a square on Earth will be a rectangle, and the El Greco Earth counterpart of a sphere on earth will be an ellipse. Also, it is important to note that the stretch is a one-off and is applied at time t_1 only, thus, even when an object's orientation relative to gravity changes after t_1, it will keep its shape. So a rectangle on El Greco Earth that is the counterpart of a square on Earth will keep its shape even if it is tipped over on its side at t_2. And an ellipse on El Greco Earth that is the counterpart of a sphere on Earth will remain elliptical even if it is tipped over so that its major axis is parallel to the ground.

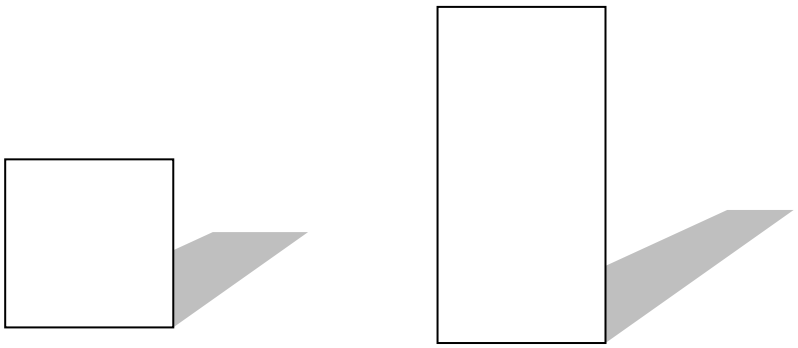
Time: t_0



Earth

El Greco Earth

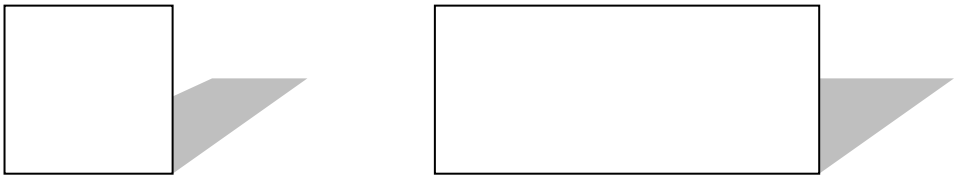
Time: t_1



Earth

El Greco Earth

Time: t_2



Earth

El Greco Earth

So, given Dom and Twin Dom acting on Earth and El Greco Earth respectively is it possible for internal states to be duplicated? Hurley argues that they cannot be duplicated unless we tamper with processes like memory, in which case we may need to admit that duplication is impossible in this case or perhaps try and shift the boundary further inwards, which has its own problems.

Consider that at time t_1 Dom on Earth is facing three spheres arranged in front of him such that the centre of each sphere is horizontally equidistant

from where he stands. On El Greco Earth at t_1, then, Twin Dom will be facing three ellipses such that the centre of each is horizontally equidistant from where he stands (the ellipses are standing on end with their major axes perpendicular to the ground). The ellipses on El Greco Earth will be twice the height of the spheres on Earth but given that the stretch has also been applied to Twin Dom's body, the heights of the ellipses relative to Twin Dom's body are of the same ratio as the heights of the spheres on Earth relative to Dom's body. If sphere one on Earth is at the same height as Dom's eyes, then ellipse one on Twin Earth will be at the same height as Twin Dom's eyes. Likewise, if sphere two is three times the height of Dom's eyes then ellipse two will be three times the height of Twin Dom's eyes and so on for sphere three and ellipse three, which are both eight times the height of the eyes of the respective viewer.

Next, suppose that on top of each sphere on Earth there is a red dot such that the line from Dom's eyes to the top of sphere one is horizontal. This will also be the case for Twin Dom and ellipse one. That is, the angle of the line from Dom's eyes to sphere one on Earth is equal to the angle of the line from Twin Dom's eyes to ellipse one on El Greco Earth. However, with regard to spheres two and three and ellipses two and three, this will not be the case. In fact, the angle of the line from Twin Dom's eyes to the red dot will be steeper than the angle of the line from Dom's eyes to the red dot in each of these cases. Furthermore, the percentage difference between the angles in each case will be different. That is, when Twin Dom is looking at the red dot on ellipse two there is a roughly 100% increase in angle compared to when Dom looking at the red dot on sphere two; for sphere three and ellipse three the percentage difference is roughly 70% (see Hurley 1998b: 316).

If we now suppose that Dom points to each red dot in turn then Twin Dom must do the same. Can internal physical states be duplicated in

these cases and if so, how? Let's look, first of all, at visual perception. The angle of incidence of light entering the retina for Dom and Twin Dom will clearly differ, resulting in different inputs to the CNS and, consequently different internal physical states, thus, initially, duplication does not seem to be possible. In order to achieve duplication in this case, Hurley suggests that: "The angle of incidence of light coming into Twin Dom's eyes needs to be transformed in a way that is a function of the twin-relative heights of the red spots" (Ibid: 316). The question here is whether or not some transformation that is made to visual input, outside of the boundary of the CNS can preserve duplication. Hurley is pessimistic with regard to this possibility but even if such a transformation, that doesn't impinge on internal states and processes, is possible duplicationists face similar problems in attempting to transform motor and proprioceptive inputs and reconcile them with the alterations to visual input. And these problems are compounded once Dom performs further actions.

Let's suppose that at t_2 Dom rolls his spheres sideways. Twin Dom will, then, roll his ellipses onto their sides such that their orientation is changed and their major axes are now parallel rather than perpendicular to the ground. If we again suppose that both Dom and Twin Dom are pointing at each of the red spots on their spheres and ellipses in turn, then we will not want a correction applied to Twin Dom's visual input in the way that it was applied when the ellipses were standing on end. As the girth of the objects did not alter when the one-off vertical stretch was applied the red dots will now be at the same height on El Greco Earth as they are on Earth. Instead, the stretch that was applied at t_1 now makes for horizontal distortions with regard to the location of the red dots on El Greco Earth relative to Earth and consequently a different kind of correction device is required in order to preserve duplication. However, we do not want this correction device to apply to all objects on El Greco

Earth as there are objects whose orientation since t_1 have not changed and that are not horizontally distorted. If this specification for the correction device is not included then the possibly of duplication in this case will be foiled by what Hurley calls 'the search –and-replace fallacy'.

There is a search-and-replace function on word processors that enables you to locate in your document every instance of the word 'red', for example, and then change it to 'green'. However, if you subsequently decide that the change should not have been made the search-and-replace function, as well as changing every instance of 'green' that had been 'red' back to 'red' again, will now change every instance of 'green' that had not previously been 'red' into 'red'. It is prudent then to tag pre-existing instances of 'green' before using this function in case you wish the change to be reversed. What Hurley calls the 'search-and-replace fallacy' throws up difficulties for the general duplication assumption. If we examine a Twin Earth in which all things that are red on Earth are, instead, green but in which not all things that are green on Earth are red, then the earthly information about the red/green distinction is lost. The information about the red/green distinction on Earth is not transformed into the information about the green/red distinction on Twin Earth. In this case a reversing device would be of no use as the distinction that it would need in order to duplicate Earthly internal physical states by cancelling out the reversing effects is not available on Twin Earth. In cases such as these, where the differences between the twin possible worlds are not symmetrical, duplication is problematic: "Mad scientists in the duplication business may be out of work for a wide range of non-symmetrical environmental differences" (Ibid: 302).

In the El Greco Earth case currently under consideration, unless the correction device can avoid compensating for objects that have not changed their orientations after t_1 then duplication will not be possible.

Given this, the correction that is applied to Twin Dom at t_2 must be sensitive to the distinction between objects that have changed their orientation after t_1 and those that have not, and will have to involve memory of some sort. Memory, on the internalist view, involves processes that occur within the CNS so, in this case the duplicationist could propose retreating further inwards in order to preserve duplication of internal physical states that do not involve memory of the objects on El Greco Earth that have changed their orientation after t_1—this is a strategy that sounds like preserving duplication for duplication's sake. Alternatively, the duplicationist could try to specify a device that can carry out the necessary transformations to visual, motor and proprioceptive inputs and can compensate for the altered relations between them, that is located outside of the CNS but within the skin, and that can track the changes in the orientations of objects on El Greco Earth that have occurred after t_1. Such a device may well be able to duplicate Dom's internal physical states in Twin Dom, but leads us to consider whether or not the spirit of the general duplication assumption is being adhered to.

The problem we are faced with is as follows: the General Duplication Assumption, as stated by Hurley, is the assumption that the duplication of *internal* physical states is unproblematic; that whatever lies inside the boundary of skin and skull should be duplicable in radically altered environments. But the duplicability of these states and processes within the skin and skull is shown to be problematic in all manner of circumstances. This led us to consider possible strategies for preserving duplication in these problem cases. The first involved shrinking the boundary around which internal physical states are to be duplicated. The first move was from the skin to the CNS in order to preserve duplicability where normal action by an agent on Earth amounted to tampering with an external reversal device on Twin Earth. But shrinking the boundary in

this way looks like the beginning of a move that is motivated merely by a desire to preserve duplication. What if similar problem issues arise for the states and processes of the CNS as in the El Greco World case above? Would the duplicationist be tempted to shrink the boundary to the frontal lobes, or some other area of the brain, such that duplication is preserved? Would it not be better, instead, asks Hurley, to question whether or not the possibility of duplication holds in general?

Alternatively, it could be proposed that the boundary around the central processes is fixed by whatever processes can be duplicated in each particular case. But if this were so, says Hurley, the boundary would be relativized and may vary from case to case. Furthermore, in certain cases we may find a certain slice of processes that are duplicated but that are not involved in any way with the adaptation to the inverse environment. Putting a boundary around these processes would not be philosophically interesting and would only be motivated by a desire to preserve duplication. One could, instead, stipulate that the boundary goes around whatever bodily and brain states can be duplicated for *any* environmental inversion. Thus, the boundary postulated by the duplicationist here would be around the intersection of the states that can be duplicated in each case. But this falls prey to the same objections mentioned above. First of all there is no reason to assume that there will be *any* processes in the intersection of the sets of duplicated states. And secondly, even if there were, it there wouldn't be any reason to regard these processes as philosophically interesting.

The second move considered above also seems to be motivated solely by a desire to preserve duplication. Instead of retreating inwards in order to preserve duplication, the boundary of the CNS is held fast and any reversing device required to preserve duplication is positioned outside of this boundary. In order to duplicate physical states and processes within

the CNS between Dom and Twin Dom what is proposed is a device that can compensate for the changes that have occurred on El Greco Earth by transforming the inputs from visual perception, motor function and propioception along with the relations between them while simultaneously tracking the objects on El Greco Earth that have (or have not) changed their orientation since t_1. All of this processing must occur outside of the boundary of the CNS if internal physical states are to be duplicated. But when we look at the internal physical states that are duplicated in this case what we find is that the duplication of internal processes that are not involved in any way with the adaptation to the altered environment. So, as with the strategy of retreating further inwards, Hurley says that there is no reason to suppose that the duplicated internal physical states will be philosophically interesting. At which point Hurley poses the question as to whether or not there is any philosophical merit to the postulation of a boundary within which the physical states of an agent should be duplicable at all.

What is it about the physical processes located within the boundary of skin and skull or within the boundary of the CNS that marks them off as being duplicable in radically altered environment? On the Internalist view of the nature of mind internal physical processes are the *central* processes between perception and action that fix the contents of mental states. Input comes to the mind in the form of perception and output comes from the mind in the form of action and the processing that occurs in between is what the mind does. The mind changes the input: perception, into output: action. The states and processes that perform this operation, on this view, are contained within the head; these are the 'central' states and processes that it should be possible to duplicate in a radically altered environment if the Internalist picture of mind is correct.

But surely, says Hurley, it is what these processes *do* rather than where they are located that makes them *central* in this regard. It is the function of the processes contained within the skin, or skull, or CNS, and not their location that bestows 'centrality' on these processes. For example, if a scientist builds and trains a neural network to replace the memory of a patient who lost or damaged a portion of his brain in an accident and attaches the neural net outside of the patient's skull then it would still be a 'central' process on the traditional view of what a central process is supposed to do. Its location is irrelevant to its ability to fulfil its function. Central states and processes should be specified functionally, then, and not according to location, says Hurley.

Hurley points out that the possibility of such a device preserving duplication of central states depends on what we mean by 'central'. She asks us to suppose that a mad scientist has built and trained an artificial neural network to perform just the comparisons and alterations mentioned in the El Greco though experiment above. Now, if the scientist inserts this artificial neural network and connects it inside the membranes that surround Twin Dom's CNS then, clearly, the physical and functional states within this boundary have been altered and duplication of Dom's central states in Twin Dom will not be possible. Thus, for this internalist interpretation of 'central' duplication fails. But what if, instead of inserting the device within the meninges, the mad scientist inserts and connects the artificial neural net in Twin Dom's elbow? In this case there will be no physical change to Dom's physical states within the CNS. Should the network still count as central? Doesn't the philosophically correct understanding of 'central' here refer to the *function* performed by the network rather than its location?[27]

[27] A la Clark and Chalmers' Parity Principle.

> So the network itself constitutes an alteration in Twin Dom's central states. Sheer location relative to an internal bodily membrane cannot provide a significant criterion of centrality for philosophical purposes. (Ibid: 318)

If, by contrast, we allow that external or extended states and processes can be 'central' then the duplication of central states may indeed be possible in radically altered environments.

Extended Vehicles

If the General Duplication Assumption is correct then the duplication of internal states and processes *should*, in general, be possible. As we have seen this is not the case. But there is something correct in the intuition that lies behind the General Duplication Assumption: if all of the subpersonal states and processes that are doing the differentially token explanatory work with regard to a particular content token are identified then those subpersonal states and processes should, in principle, be duplicable in radically altered environments without a problem. That is, once you have identified the entire vehicle of content then that vehicle should be duplicable. As Hurley puts it: "If a vehicle, then duplicable" (Ibid: 331). So, if we have identified the complete vehicle that carries a particular content then this vehicle should be duplicable in different environments. However, if the subpersonal states and processes are not duplicable in different environments then you have failed to identify all of the subpersonal states and processes doing the token-explanatory work. Again, as Hurley puts it: "if not duplicable, then not the whole vehicle" (Ibid.).

Hurley suggests, then, that the duplicationist has not identified *all* of the subpersonal states and processes that are doing the differentially token

explanatory work. Given that the internal states and processes that the duplicationist wishes to duplicate are not always duplicable, this suggests that the duplicationist is failing to identify the whole of the vehicle by focusing exclusively on *internal* states and processes. In order to identify all of the subpersonal states and processes that are doing the token explanatory work, then, we must look externally, says Hurley. We must look externally for the remainder of the differentially token explanatory subpersonal states and processes; we must look externally in order to find the whole vehicle in certain cases.

The reason, for Hurley, that the General Duplication Assumption does not hold is that it is predicated on the faulty assumption of Internalism about vehicles. Hurley says that the reason the duplicationist runs into problems in the cases identified above is because he has assumed that the differentially token explanatory states and processes of particular content tokens must be *internal* states and processes of the organism. He will, therefore, fail to identify the whole vehicle in certain cases because he is constrained by his default starting position of Internalism. If we abandon the assumption of Internalism, then, and allow that the vehicles of content can in some cases go external then it should be possible to achieve duplicability of subpersonal token explanatory states and processes in all cases. To quote Hurley once more: "if duplicable, then a vehicle, and if not the whole vehicle, then not duplicable" (Ibid.). Duplication of the states and processes that realise token mental contents is possible then but the General Duplication Assumption does not hold as it maintains that these states and processes must be internal. Duplication is possible but only provided that the whole vehicle is identified, if we fail to identify the whole vehicle because we are constrained by the assumption of Internalism, for example, then duplication will not be possible in some cases.

If we return to the Twin Otto thought experiment one last time then we can see that there is good reason why Twin Otto cannot be a physical duplicate of Otto from the skin in. In this case we have a counterfactually altered environment—one where Twin Otto writes '51' instead of '53'. Duplication of internal physical states in this case requires the duplication of information in the notebook. Clark and Chalmers failed to identify the whole of the vehicle in this case. If Hurley is correct then the vehicle of content for Otto's belief *that the Museum of Modern Art is located on 53rd Street* includes the notebook and what is written in it. Similarly, the vehicle for Twin Otto's belief *that the Museum of Modern Art is located on 51st Street* includes his notebook and the mistakenly entered '51'. Otto and Twin Otto then are not twins, they are not physical duplicates from the skin in and the vehicle of content for their respective beliefs about the Museum of Modern Art differs also. This is why Otto's belief about the Museum is true and Twin Otto's belief is false.

Constitution vs. Enabling

What I take Hurley's arguments to demonstrate, with great skill and an amazing thoroughness, of which I hope to have given just a tiny flavour here, is that the assumption of general duplication is a mistake; that it is not, in general, possible to duplicate internal physical states in counterfactually altered environments. In doing so she also demonstrates that to specify the boundary of skin and skull as a 'magical membrane' outside of which enablers of content and cognition cannot fall is a mistake and that the vehicles of content can indeed extend beyond the boundaries of skin and skull.

But we should be wary of supposing that appeal to the spread of vehicles between organism and environment should provide a basis for EMT. If Hurley's take on the status of vehicles, is that they are *enablers* of mental

states rather than constitutive of mental states then the extension of mental states required for EMT does not follow from the extension of vehicles. Hurley frequently interchanges 'vehicle externalism' with 'enabling externalism' and refers to vehicles themselves as enablers of content:

> They explain, e.g., what processes or mechanisms enable a given intention to look inside the box on the left, or a given visual experience of a certain surface as blue. (Hurley, forthcoming)

Vehicles, as enablers of particular mental states, cross back and forth between organism and environment. But this does not entail extension of the mental states themselves. Additionally, in emphasising the vehicle content/distinction as strongly as she does Hurley highlights that an enabler of a mental content or mental state may not be constitutive of that mental state. Thus, if the extension of mental states is required for EMT then Hurley's Vehicle Externalism is not EMT—it is EMMT. All of which leaves the door open for someone who wishes to reject EMT and remain internalist about mental states while yet endorsing vehicle externalism, EMMT.

So, if the extension of mental states is required for a genuine EMT and if Hurley's vehicle externalism is really only a form of enabling externalism that secures the extension of the enablers of mental states, or the machinery of mind, then Hurley, like Rowlands, does not argue for a genuine EMT in the way that Clark and Chalmers do. But, as we saw in chapter one, Clark and Chalmers' arguments fail to definitively secure the extension of mental states into the world. Part of the reason that they have failed in this may be that they are working without a definitive mark of the mental. This is an issue that we touched on briefly at the end of

chapter 2 and I take it up further in the next chapter as we examine the most prominent objections that have been levelled at EMT.

Chapter 4

Introduction

In this chapter I consider some of the objections that have been levelled at EMT, doing so will reveal that objectors typically target Clark and Chalmers' claims for the extension of cognitive processes rather than mental states into the world. I argue that that all of the serious objections considered reduce to the constitution versus enabling issue. With regard to cognitive processes this issue can be settled by the provision of an independently motivated mark of the cognitive. And with regard to mental states the cleanest way to settle the issue is by the provision of an independently motivated mark of the mental. Examination of the objections will involve consideration of arguments against the extension of cognitive processes as well as against the extension of mental states, but it should be borne in mind throughout that the extension of mental states is required for a genuine EMT.

Not all cognition is extended

The first objection I consider is one based on a misunderstanding of the claims of EMT and for this reason it is rarely found in the literature. The objection in its most basic form is that not every cognitive process is an extended cognitive process. In a slightly more sophisticated form the objection may be that in certain cases the use of external resources to aid in the completion of a cognitive task is not necessary and in fact may be a hindrance rather than a help, for example, and it is along these lines that Rupert (2004) objects to Rowlands' (1999) characterisation of memory.

As we saw in chapter two, Rowlands (1999) argues that the type of memory that is most characteristic of modern literate human beings is

prone, in the right circumstances, to extend into the environment. He claims that the advent of external means of representation means that it is practically possible to reduce our reliance on our internal memories in certain circumstances, transforming the nature of the cognitive task facing the individual. In such cases the external information-bearing structures play an essential role in the completion of the cognitive task and, according to Rowlands, constitute a proper part of the cognitive process.

Rupert makes a number of interesting points against this view, chief of which is his claim that Rowlands does not do enough to prove that his arguments demonstrate a fully fledged EMT, which if true would entail the literal extension of mental states and cognitive processes into the world, rather than the less radical enabling externalism, which I characterise in chapters one to three as EMMT: the Extended Machinery of Mind Hypothesis. Indeed, Rupert argues that what he calls the Hypothesis of Embedded Cognition (HEMC) offers a superior explanatory paradigm than EMT (or HEC, the Hypothesis of Extended Cognition as Rupert calls it). HEMC is characterised as follows:

> According to the hypothesis of embedded cognition…cognitive processes depend *very* heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place. (Rupert 2004: 393)

Seeing cognition as embedded rather than extended allows that the enablers cognitive states can extend beyond the boundaries of the organism and into the environment, while stopping short of the claim that cognition itself or cognitive states, extend into the environment. In fact, as we will see in the course of this chapter, many objectors take EMMT

to be uncontroversial and relatively unproblematic in the same way as Rupert does. The target for most objectors is, therefore, EMT rather than EMMT. And Rupert argues that Rowlands gives us no reason to endorse EMT (HEC) over the less radical EMMT (HEMC).

> Rowlands, however, does not make clear why the use of an internally represented code applied to the contents of an external store implies HEC, rather than HEMC. Although increased use of external resources might change the character of internal processing and the way in which the subject interacts with her environment, why think that the apposite external and internal states (or forms of processing) are thereby of the same causal-explanatory kind? Why infer the existence of one overarching kind, memory, subsuming both internal and external states and processes that will be of significant explanatory use in cognitive science? (Ibid: 408-409)

This is a legitimate criticism and one that we will look at in more detail in later sections. For now, however, I want to focus on a different worry that Rupert articulates. Rowlands, following Miller (1956), highlights the limitations of human biological working memory and argues that we need to look beyond biology to the external world in order explain our proficiency at certain memory tasks. Rupert concedes the point regarding the informational limitations on biological working memory but argues that Rowlands' EMT presents a view of biological memory that is not consonant with the research:

> Although we must recognize limitations on the capacity of working memory, Rowlands' view fits poorly with much of the empirical data. Consider the human ability to converse effectively. Participating in a conversation of any significant length makes rigorous demands on working memory by requiring participants to build and maintain a fairly detailed model of the ongoing discourse. (Rupert 2004: 409-410)

Rupert goes on to illustrate the difficulties in explaining how we conduct a normal everyday conversation along what he thinks are EM theorist's lines. He asks us to imagine attempting to converse with someone while using an Otto-style working memory—maintaining a written running model of the discourse on paper—rather than biological working memory. Such a strategy would be, needless to say, prohibitive of a smooth conversation.

> In the context of a standard verbal exchange of any significant length, external resources are virtually useless, while internal storage appears to be irreplaceable. (Ibid: 410)

Thus, in this case the use of external resources is not only not required, it may well be a hindrance in facilitating the completion of the cognitive task in question. The example is supposed to show that Rowlands may have been overly hasty pointing to the limitations of biological working memory as a motivation for the use of external information-bearing processes.

To be clear, neither Rowlands, nor any other EM theorist, wholly dismisses biological working memory in favour of external information-bearing resources. On the EM model extended cognitive processes are hybrid and cannot exist without the internal portion of the process. In short, there is no extended working memory, of the sort that Rowlands claims is characteristic of modern literate human beings, without biology.

But the problem for Rupert's objection is that it mischaracterises EMT in a different way. No EM theorist expects that EMT should provide a framework within which *all* cognition is partially constituted by external information-bearing processes. It is entirely expected, on this view, that many cognitive processes—and, indeed, all biologically basic ones—will be completely internally constituted. Thus, it is no argument against EMT to demonstrate that the dynamics of conversation resist explanation along

externalist lines since EMT only holds that *some* cognitive processes are partially externally constituted, not that all are.

This point is well made by Hurley (forthcoming) in demonstrating why EMT has a lower burden of proof than internalism.

> Internalism claims to characterize all mental states, and externalism denies this claim must hold without itself claiming to characterize all mental states. Externalism thus has a lower burden of proof than internalism: externalism is vindicated by providing counter-examples to internalism, but internalism is not vindicated by providing counterexamples to externalism. Externalism can accommodate examples of internalist explanation with equanimity, since it denies that internalism's universal ambitions are justified without adopting comparable universal ambitions of its own.

Rupert's strategy of providing a counterexample to EMT by highlighting its deficiencies with regard to explaining how we can carry out conversations fails because the EM theorist can simply agree that Rupert is correct in this case while pointing out that it was never part of EMT to argue that all cognitive process require externalist explanation; EMT may well apply in many other cases of cognitive processing.

Differences

The differences objection is perhaps the most obvious one to level at EMT and I have already dealt with it to a certain extent in consideration of the integrationist perspective on the Parity Principle in chapter one. The objection is that external processes, or the external portion of extended processes, are just *too unlike* internal processes to count as mental or as cognitive. It is along these lines that Rupert makes the following statement:

> I argue that the external portions of extended ‘‘memory’’ states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds; this quells any temptation to argue for HEC from brute analogy (viz. extended cognitive states are like wholly internal ones; therefore, they are of the same explanatory cognitive kind; therefore there are extended cognitive states). (2004: 407)

Once more, Rupert is not arguing that external processes have no role to play in the enabling of certain cognitive or mental states; rather his claim is that these external processes are so unlike inner ones that they cannot legitimately be coupled with internal processes to form extended processes that are of the same cognitive kind as purely internal processes. Thus, his target is EMT rather than EMMT.[28]

He notes that highlighting the differences between putative extended cognitive processes and internal cognitive processes removes any temptation to argue from ‘brute analogy’; a strategy which is, needless to say, not a good one. Fortunately for the EM theorist this is not a strategy that gets employed. Despite the key role accorded to the Parity Principle by the EM theorist, it is not claimed that the external aspects of putative extended mental states should be ‘just like’ internal states to which we would accord mental status. Rather, as we saw in chapters one and two, the differences between internal processes and the external portion of extended processes are *predicted* and *required* by EMT.

This is the insight that is provided by the Integrationist perspective on EMT (see chapters 1 & 2, and Menary 2006; 2007, Sutton 2006). The idea is that we offload portions of our cognitive processing into the environment precisely because the structures and processes that we find

[28] Although it may be a very crude version of EMT that he targets.

in the environment are different from the internal brain processes and so afford us, for example, a stability and reliability that our brain processes do not (See Otto). As Rowlands puts it: 'without these differences the external processes would be otiose" (2009a: 4).

Nonetheless, there is still something to the differences objection proposed by Rupert. Even if the Integrationist perspective on EMT requires that the external portion of extended cognitive processes differ in some important respect from purely internal processes, extended processes (comprising internal and external aspects) must be of the same kind as purely internal processes if both are to be properly called cognitive. In the absence of a specification of what it is in virtue of which extended processes are of the same kind as internal ones, the option is open for those hostile to EMT to reject it on the basis that the differences are too great. Thus, objectors to EMT can argue that *even if EMT requires a difference* between internal states and processes and the external portion of extended states and processes, the differences between the two are so great that they cannot be of the same kind, and so extended states and processes are neither cognitive nor mental.

It is in this vein that Rupert argues that what the EM theorist refers to as extended memory differs too greatly from what we already know about memory to be considered cognitive. For example, Rupert cites a study that examines the idea of negative transfer with regard to memorising sets of pairs. Negative transfer is a particular instance of what are generally referred to as interference effects; effects that inhibit an individual's capacity to learn and remember new information. Suppose a subject is presented with a list of words that he must commit to memory, the list consists of paired male and female names such as John and Sally. We can call this list the 'A-B' list. When the list has been memorised the experimenter presents the subject with one of the 'A-words' from the list

and it is the subject's task to utter the correct 'B-word', which is the word that it had been paired with on the list, i.e. if the experimenter says 'John', the subject should respond with 'Sally'. In the recall task then, the 'A-word' is the stimulus and the 'B-word' is the response. Following this task the experimenter mixes the associations that appeared on the 'A-B' list such that John is now paired with Mary instead of Sally, who is now paired with Steve. Call this new list of pairs the 'A-C' list. The subject proceeds as before to learn the new 'A-C' list of paired associations. But results indicate that subjects learn these new associations a lot more slowly than the initial 'A-B' list. The reason for this is outlined by Rupert as follows:

> There is, it is said, negative transfer, an interference of the old associations with the learning of the new. The problem seems to be that if, for instance, John was married to Sally according to the A-B list, subjects have a hard time blocking out this association and forming a new association between 'John' and, say, 'Mary', with which 'John' is now paired on the A-C list. (Rupert 2004: 416)

Rupert next asks us to consider the case of an individual who uses a notebook and a pen to record the 'A-B' and 'A-C' lists as in Clark and Chalmers' Otto case. Rupert says that there is unlikely to be any negative transfer in cases such as this as all the subject needs to do is read the paired associations from the written lists in each case. The EM theorist could respond that the stability provided by the notebook is the reason that no negative transfer occurs in cases such as this and it is precisely for these sorts of reasons that external resources are recruited for the completion of cognitive tasks. But such a response is unlikely to cut any mustard with Rupert who holds that the differences between internal memory processes and putative extended memory processes that he has highlighted "are at the very heart of cognitive scientists' investigations of

memory" (Ibid: 415). For Rupert, the use of notebooks and other external artefacts for the completion of cognitive tasks is best seen as an example of *embedded* cognition rather than as an example of extended cognition precisely because the use of artefacts is merely an example of the use of memory *aids* that are peripheral to the real process of remembering; a process that exhibits negative transfer and the generation effect for example. Rupert's specification of genuine memory as being required to exhibit negative transfer and the generation effect is an attempt to specify a mark of the cognitive for genuine memory; a criterion, or criteria, that distinguish(es) cases of genuine remembering from cases of the use of remembering aids, for example.

Rupert may well be correct in his analysis of these psychological laws which seem to be uniquely characteristic of internal memory processes but it is not clear that they provide a sufficient basis for delineating the cognitive from the non-cognitive, which is what is required from a mark of the cognitive. Bartlett (2008) notes that features such as these are *contingent* features of human psychology and points out that certain savants have memory systems that seem to work very differently from those of so-called normal individuals but we do not, on that basis, deny that they are remembering. Similarly, the commonly referenced '7 units (plus or minus 2 units)' rule (see Miller 1956) for information processing is seen as typical of the capacity of human short-term memory and is one that putative extended memory systems recruiting external artefacts are unlikely to be limited by in the same way as basic biological memory. But it is not clear that we should withdraw cognitive status from internal processes associated with short-term memory if this rule is not adhered to, particularly if it is exceeded.

Along the same line of objection Adams and Aizawa protest that:

> Cognitive Processes are so different from the physical processes in the tools we use that a science that ignores this difference essentially ignores cognition. (Adams & Aizawa 2001: 47)

There is a danger evident in this quote of doing a disservice to EMT by characterising it as the *External* Mind Thesis rather than the Extended Mind Thesis. Adams and Aizawa frequently (though not always) flirt with this mischaracterisation of EMT by comparing and contrasting internal processes with external processes instead of internal processes with extended processes, which are identified as hybrid entities comprising internal and external processes. This is not to say that the external portion of the process is not a legitimate target for objectors to EMT, just that it needs to be stressed that the EM theorist will emphasise that the external process is just part of the larger process that necessarily includes internal processes and that to ignore this fact is to be in danger of attacking a straw man.[29] Bearing this is in mind I proceed to examine Adams and Aizawa's version of the differences objection.

In relation to Clark and Chalmers' Tetris cases that we considered in chapter one, Adams and Aizawa note that the process of mentally rotating the images of blocks is very different from rotating images of blocks on a computer screen by pressing a button.

> It seems to us safe to assume that the process that physically rotates the image on the screen at the push of the button… is not the same as the cognitive process that occurs in the brain. Pushing the button closes some sort of electrical circuit that, at some extremely short time delay, changes the way electrons are fired at the phosphorescent screen of a cathode ray tube. This sort of causal process is surely not the same as

[29] As we will see, this strategy is crucial for the EM theorist in responding to Adams andd Aizawa's (2001) mark of the mental objection.

> any cognitive process, or any fragment of a causal process, in the brain. (Ibid: 54)

Adams and Aizawa contrast internal processes with external processes here; a strategy that, as noted above, does not do justice to the EMT position. Simply pointing out these obvious differences is not going to provide a significant challenge to the EM theorist. The EM theorist will argue that what matters in cases such as this is that the internal processing in the mental rotation case and the extended processing involving the rotation of the image on the computer screen both fulfil the same functional role by contributing in the same way to the completion of the cognitive task.

To be fair to Adams and Aizawa they go on to highlight differences between the internal processing involved in a case of purely mental rotation and the internal processing involved in a case of on-screen rotation. But again here the EM theorist can simply re-emphasise the Integrationist element of EMT whereby precisely these differences are predicted and required. Thus, once more it seems that appeal to these sorts of differences will not do the work required by EMT's detractors.

It is not enough to point to some difference between internal processes and external or extended processes and deny cognitive or mental status to the extended process on the basis that the whole process is not *enough like* a genuine cognitive process or mental state. Clearly, external processes play an important role in both the Otto case and the Tetris case but the challenge for both sides is to provide an account of which states and processes are genuinely constitutive of cognition and which states and processes are merely *enablers* of cognition. Likewise, in order to settle the debate over the stronger claim regarding the extension of mental states, the specification of a criterion that clearly delineates the

mental from the non-mental; what is constitutive of a mental state from what merely enables mental state formation is required.

The differences objection, then, leads us to focus on the big issue of constitution versus enabling that I highlighted in chapters one to three. It is not enough to point out the differences between internal processes and external processes in order to deny cognitive status to extended processes. One should first account for what a cognitive process is and then adjudicate on boundary issues.

Likewise, it is not enough for the EM theorist to point out similarities between internal processes and extended processes in order to confer cognitive status to the latter: the requirement for an account of what distinguishes a process as cognitive applies equally here. So, the issue can only be settled by the provision of an independently motivated account of what exactly makes a process cognitive, or what makes a state mental, regardless of its location. What is required from both sides of the debate over the boundaries of cognition then is a mark of cognition.[30]

If EMT requires the extension of mental states and not just the extension of cognitive processes then even an independently motivated mark of cognition will not move the debate forward with regard to the metaphysical extent of minds. What is needed to settle the issue here is the provision of a mark of the mental.

The Coupling/Constitution Fallacy

According to Adams and Aizawa (2001, forthcoming) many arguments presented in favour of EMT are fallacious because they confuse instances

[30] To give Adams and Aizawa their due once again, they do recognise this issue and acknowledge that their differences objection forms part of a larger argument that incorporates a Mark of the Cognitive. This is something I explore in a later section.

of coupling with cases of constitution. Once again, EMMT is taken to be unproblematic by the objectors, which is to say that it is acknowledged that external factors do have a role to play in the completion of certain cognitive tasks. However, the point of contention is the movement from this innocuous EMMT to a fully fledged EMT. It is precisely this step that is at issue in the coupling/constitution debate.

> Coupling arguments are far and away the primary sort of argument given in support of transcranialism. What is common to these arguments is a tacit move from the observation that process X is in some way causally connected (coupled) to a cognitive process Y to the conclusion that X is part of the cognitive process Y. The pattern of reasoning here involves moving from the observation that process X is in some way causally connected (coupled) to a process Y of type Φ to the conclusion that X is part of a process of type Φ. In attributing this pattern of reasoning to advocates of transcranialism, we do not mean that they consciously and deliberately draw a distinction between the coupling claim and the constitution claim, and then explicitly assert that coupling is sufficient for constitution. Far from it. What typically happens is that writers just casually slip between one and the other. (Adams and Aizawa, forthcoming)

The implication is that EM theorists see coupling relations between a cognitive process and an external process and get overexcited, positing constitution where there is none. But why, exactly, is the move from causal coupling to constitution a fallacy? To illustrate the point Adams and Aizawa offer us the example of a bi-metallic strip in a thermostat for consideration. The expansion of a bi-metallic strip in a thermostat is:

> …[C]ausally linked to a heater or to an air conditioner that regulates the temperature of the room the thermostat is in. Expansion does not, thereby, become a process that extends to the whole system. It is still

> restricted to the bi-metallic strip in the thermostat. (Adams and Aizawa 2001: 56).

In this case to assume that expansion is a process that extends to the rest of the thermostat system simply because the rest of the thermostat is causally coupled to the bi-metallic strip would be a mistake. Mere causal coupling does not entail that the coupled object or process becomes partially constitutive of that to which it is coupled. Similarly, Adams and Aizawa argue that just because an external object or process is causally coupled to a cognitive process in the brain does not mean that that object or process now becomes constitutive of that cognitive process. In making arguments of this type, the EM theorist is guilty of confusing coupling relations and constitution according to Adams and Aizawa.

It could be argued here that the example chosen by Adams and Aizawa does not do justice to the complexity of EMT. Of course the expansion of the bi-metallic strip does not extend to the rest of the thermostat—it is expansion *of the bi-metallic strip*! The claim of the EM theorist is not that neural firings in the brain extend into external processes and objects. It is that cognition extends and the opportunity to make this claim exists because (unless mind-brain type-identity theory is correct) cognition is not brain-bound by definition, unlike neural firings.

Anyhow, it remains the case that causal coupling between two processes does not necessarily licence the claim that both processes form a single, cognitive kind. Thus, Adams and Aizawa legitimately criticize Wilson (2004) for arguing that in the strategies we employ in solving certain puzzles the mind: "extends itself beyond the purely internal capacities of the brain by engaging with, exploiting, and manipulating parts of its structured environment" (Ibid.: 195). Consider the attempt to form as long a word as possible using single letters marked on a random

assortment of individual tiles, as in the word-game Scrabble. The typical way of completing this task is to arrange and re-arrange the tiles such that different combinations of letters present themselves. The idea is that manually re-arranging the tiles in this way is more fruitful and efficient than simply looking at the static tiles and trying to generate novel combinations 'in the head'. The more typical strategy is an example of an epistemic action of the sort explored by Kirsch and Maglio that we considered in chapter one. In cases such as these Wilson argues that epistemic actions are constitutive of cognition:

> We solve the problem by continually looking back to the board and trying to figure out sequences of moves that will get us closer to our goal, all the time exploiting the structure of the environment through continual interaction with it. We look, we think, we move. But the thinking, the cognitive part of solving the problem, is not squirreled away inside us, wedged between the looking and the moving, but developed and made possible through these interactions with the board (Ibid: 194).

Adams and Aizawa take Wilson to be arguing for the literal extension of cognition, of thinking, into environment in our interactions with the scrabble tiles or puzzle board. They claim that if this is indeed what he means, the literal extension of cognition rather than of the processes supporting cognition, then he is guilty of the coupling/constitution fallacy. Note, once more, a lack of clarity here in the claims of the EM theorist—we can contrast the early part of the last sentence of Wilson's quote: "the thinking, the cognitive part of solving the problem, is not squirreled away inside us…" (Ibid: 194), which is a clear expression of partial external location/constitution of cognition, with the latter part of the sentence where he describes thought as being: "developed and made possible through these interactions with the board" (Ibid: 194), which is

an expression of EMMT. If EMMT is all that Wilson is arguing for then Adams and Aizawa emphasise that they have no problem with his claim as he is "providing no argument for transcranial cognition." (Adams & Aizawa, forthcoming). But if he is indeed arguing for EMT then he is guilty, they claim, of the coupling/constitution fallacy and has failed to provide a case for transcranial cognition. The same charge can be levelled at Clark and Chalmers' Otto case on this view.

Adams and Aizawa also note a tendency of EM theorists to argue for the extension of cognition from a systems perspective, which is to say that the EM theorist points out that the individual and the external information-bearing structure to which he is coupled forms a cognitive system. It is a cognitive system because it is a system that is formed precisely for the purpose of completing a cognitive task (e.g. Otto and his notebook) and because the external portion of the system has an essential role to play in the completion of the cognitive task; thus cognition is said to include this external portion of the system.

Adams and Aizawa are willing to grant that the individual in conjunction with the external information-bearing structure to which he is coupled can form a system. They are even willing to concede that this coupling forms a cognitive system. But they correctly point out that it is one thing to claim that the coupling of two objects or processes forms a cognitive system and quite another thing to claim that cognition extends to every part of that system.

> We can...concede that humans and their tools constitute cognitive systems. Still, this does not establish transcranialism. It does not follow from the fact that one has an "X system" that every component of the system does X. Obviously there are systems that

> consist of many types of components and involve a multiplicity of processes. (Ibid.).

To illustrate their point here Adams and Aizawa (forthcoming) ask us to consider a computer. Not every process that occurs in the computer will be a computational process, not every process that occurs as part of this individual computing system will be a process that computes. For example, a fan causes cool air to circulate, mitigating the effect of the heat produced by the CPU. So, say Adams and Aizawa, appeal to the notion of a cognitive system is not likely to help the EM theorist and the charge of confusing coupling relations with constitution can still be laid at their door.

However, as Rowlands points out it is far from clear that the EM theorist is simply *confusing* coupling with constitution.

> Far from confusing constitution and causal coupling, the most natural way of understanding the arguments for EM are precisely as arguments for reinterpreting what had traditionally been regarded as extraneous causal accompaniments to cognition as, in fact, part of cognition itself. And, in general, to argue for the identification of X and Y, when X and Y had hitherto been regarded as distinct types, is not to confuse X and Y. (2009a: 5).

This is correct, after all, not every instance of causal coupling is argued to be a case of extended cognitive processing and in those cases where coupling is argued to lead to the partial external constitution of a cognitive process it is not *mere* casual coupling that is said to occur. Rather, extended cognition is said to occur when external factors play a special enabling role in the completion of some cognitive task. EM theorists, as Hurley puts, make a *discriminating* appeal to causal spread (hence Clark's list of additional criteria that we examined in chapter one).

One can argue that they do not succeed in making this discriminating appeal, and one can highlight loose talk that slips too easily between EMT and EMMT. Still, to indict the EM theorist for *confusing* coupling relations with constitution is a different charge and one that isn't warranted.

However, although a proper appreciation of the strategy of the EM theorist in this regard means rejecting charges of simply *confusing* causal coupling with constitution, it is not clear that the discriminating appeal to causal spread—in the form of the additional criteria offered by Clark that we saw in chapter one—will ease worries over the issue of constitution versus enabling. The cleanest way to settle this issue, as we will see is by providing a mark of the mental. Next, however, I examine another objection from Adams and Aizawa that trades on worries about causal spread.

Cognitive Bloat

> If Clark and Chalmers opt for the simplistic view that anything that is causally connected to a cognitive process is part of the cognitive process, then there is the threat of cognition bleeding into everything. This is sometimes called something like "the problem of cognitive bloat" or "cognitive ooze". These names do justice to the ugliness of the view, but not to its radical nature. The threat is pancognitivism, where everything is cognitive. (Adams & Aizawa 2001: 57).

The Cognitive Bloat objection is a slippery-slope type objection (also noted by Rowlands (2009a)) that is supposed to highlight the dangers of allowing structures and processes that are external to the agent, such as the sentences in Otto's notebook, to count as partially constitutive of cognitive processes or mental states. If we allow these sorts of states and

processes to be partially constitutive of cognition or mentality then, the objection goes, what is to stop us allowing the numbers in phonebooks or search results from Google to count as partially constitutive of cognitive processes or mental states?

Once more, the objection is not about whether or not external objects or structures such as notebooks have a role to play in the completion of certain cognitive tasks (EMMT), rather, the objection concerns whether or not the role that these objects and structures play in the completion of cognitive tasks warrants attributing cognitive status to these structures and objects as constitutive parts of the cognitive process. So, the objection is that allowing EMT puts one on a dangerous slippery-slope whereby all manner of things that are patently non-cognitive must be admitted to be cognitive—EMT forms, on this view, a *reductio ad absurdum*.

Interestingly, Fodor (2009) claims that Clark and Chalmers' argument in favour of EMT is actually a slippery-slope argument itself and that the much vaunted Parity Principle is not required by EM theorists:

> [Clark's] real argument is that, barring a principled reason for distinguishing between what Otto keeps in his notebook and what Inga keeps in her head, there's a slippery slope from the one to the other. That being so, it is mere prejudice to deny that Otto's notebook is part of his mind if one grants that Inga's memories are part of hers.

Fodor goes on to point out that slippery-slope arguments are notoriously invalid but states that in any case he can meet Clark and Chalmers' challenge by providing a principled difference between Otto and Inga. We will consider this principled difference in the next section on the mark of the mental.

For now, however, we can note that in response to the cognitive bloat objection Clark introduces a series of conditions that must obtain in order for an extended process to count as cognitive, thus attempting to provide a basis for distinguishing between extended cognitive and external non-cognitive processes. So, as we saw in chapter one, Clark argues that for an external information-bearing resource, such as Otto's notebook, to count as part of a mental state or cognitive process it is necessary:

1. That the resource be reliably available and typically invoked.
2. That information contained in the resource should be easily accessible as and when required.
3. That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
4. That the information should have been previously consciously endorsed by the subject.

These conditions limit, or at least are supposed to limit, which cases of the use of external information-bearing structures legitimately count as being partially constitutive of cognitive processes. Thus, one's ordinary use of the phonebook to locate a telephone number would not typically count as an extended cognitive process because it would fall foul of conditions one, two and possibly four. Similarly, use of search engines like Google would not count as a case of cognitive extension because search results would fall foul of conditions three and four.

Kenneth Aizawa argues that these conditions are prohibitively strong in that they withhold cognitive status from obviously cognitive internal processes:

> I will argue here that these conditions are so restrictive that clear cases

> of an agent's cognitive processing will turn out to be excluded from that agent's cognitive processing. (Aizawa, unpublished)

In support of this claim Aizawa offers two cases concerning memory. In the first case we are asked to imagine Dotto who hits his head on a cabinet. Dotto suffers minor bruising but suffers no cognitive impairment. However, he fears that his memory may have been compromised by the incident and decides to reduce his dependence on it. Consequently, when asked for a phone number that he would previously have known, he replies that he does not know despite the fact that the number would "flash before his mind" (Ibid.). Thus, Dotto no longer automatically endorses (condition 3) the information obtained from his biological memory, and no longer typically invokes it (condition 1), preferring to rely on phone directories etc. So, says Aizawa, on Clark's account despite the fact that the psychological processes are the same for Dotto before and after the accident we must withdraw cognitive status from Dotto's memories as he no longer automatically endorses and typically invokes them. The second case is essentially the same as the first except in this case Dotto withdraws automatic endorsement of his output from memory because he has undertaken a study of long-term biological memory and has found that it is unreliable.

Aizawa argues that the correct interpretation of each of these cases is that Dotto has a cognitive resource that he no longer typically invokes or automatically endorses. But in each case, says Aizawa, we are prevented from applying this interpretation if Clark is correct. Instead, we are forced to say that internal psychological processes which once were cognitive processes are no longer cognitive processes. Now, whether or not this objection works depends upon what role Clark sees his additional criteria as playing with regard to cognitive status. Are his criteria offered jointly as a mark of the cognitive simpliciter, in which case they would

have to be adhered to by any putative cognitive process, internal or external? Or are these criteria offered merely as a set of additional criteria to be fulfilled by an external resource to form part of a putative extended cognitive process?

If Clark's conditions are offered as a mark of the cognitive simpliciter then Aizawa may well be right in arguing that the conditions are prohibitively strong in ruling out obviously cognitive processes, as outlined above. However, it is clear that this is not what Clark has in mind when he offers these conditions.

> In response to… concerns about availability and portability, we offered a rough-and-ready set of *additional criteria* to be met by *non-biological candidates* for inclusion into an individual's cognitive system. (Clark, forthcoming, emphasis mine)

The conditions offered by Clark are supposed to function as extra conditions on non-biological processes (presumably on top of the conditions on biological processes, which he has not given us) in order for them to be considered as part of a cognitive process. Thus, the conditions function more like a mark of the extended cognitive rather than as a mark of the cognitive. Aizawa's worries are misplaced then as the power of his objection is contingent upon Clark's conditions forming criteria for any process being cognitive.

Nonetheless, Clark's strategy in meeting the cognitive bloat objection highlights once again that the constitution versus enabling issue has not adequately been addressed. Clark has not told us exactly what it is that unites internal cognitive processes and putative extended cognitive processes as cognitive. So, although the force of the cognitive bloat objection may have been neutralised, the question of why the external

special enablers of cognition should form a constitutive part of the cognitive process at all still remains. Thus, analysis of the cognitive bloat objection leads us, once again, to the core issue of constitution versus enabling. The cleanest way to deal with this issue would be to provide a mark of the cognitive but Clark has not been forthcoming. More importantly for our purposes, settling claims regarding the extent of mental states rest on the provision of a mark of the mental we have received no indication for what this might be from Clark.

The Mark of the Mental

With the exception of the first objection, which is based on a misunderstanding of EMT, all of the objections considered to this point reduce to the constitution versus enabling issue. The EM theorist argues that in certain cases, processes that involve the manipulation of external information-bearing structures are partially constitutive of cognitive processes. However, without the provision of an independently motivated mark of the cognitive it is not clear that these claims can be definitively secured. Furthermore, without the provision of an independently motivated mark of the mental it is not clear that Clark and Chalmers can definitively secure a genuine EMT, as opposed to EMMT.

But the same demand falls on the objector to EMT. The EMT 'hostile' should provide an independently motivated mark of the mental with which he can provide a principled difference between internal mental states and putative extended mental states. Otherwise he leaves himself open to charges of question-begging. So, in this section I examine some of the objections to EMT that have developed around the requirement to provide a mark of the mental.

As noted in the previous section, Adams and Aizawa (2001) adopt a position described as *contingent intracranialism*, arguing that although transcranialism is neither a logical nor nomological impossibility, it is a matter of contingent empirical fact that all human cognition takes place within the head. The reason that Adams and Aizawa offer us for this position is that putative extended cognitive processes do not bear the mark of the cognitive. According to Adams and Aizawa cognitive processes must involve non-derived, or underived, content. Thus, Adams and Aizawa offer non-derived content as a mark of the cognitive. But, since Brentano, non-derived content is typically taken to provide a mark of the mental. It is possible that Adams and Aizawa offer it as a requirement that cognition involve processing on mental states with non-derived content. In any case, they characterise non-derived content as follows:

> Underived content arises from conditions that do not require the independent or prior existence of other content, representations or intentional agents. So, for example if minds evolved, the first mind did not acquire its thought content from any other mind (there were no others). (Adams & Aizawa 2005: 662)

Fodor (2009) also emphasises the requirement for non-derived content, but he makes it with regard to mental states rather than cognitive processes.

> [O]nly what's literally and unmetaphorically mental has content, but... if something literally and unmetaphorically has content, then either it is mental (part of a mind) or the content is 'derived' from something that is mental.

So for Fodor non-derived content is the mark of the mental, which is to say that minds and only minds possess or exhibit non-derived content.

Fodor argues that Otto's notebook cannot literally be part of his mind then, as EMT claims, because Otto's notebook does not have non-derived content. The entries in the notebook can be about something, they can exhibit intentionality but not in the same way that Otto himself can exhibit intentionality because the entries in Otto's notebook are dependent on him, on his thoughts and intentions, which exhibit their own non-derived intentionality. The contents of Otto's notebook are derived and have meaning only in virtue of the prior existence of Otto as a content-bearing, intentional agent. Similarly, words, numbers, and traffic signals in general all have content, they all mean something, and they all mean something by convention. Without the logically prior existence of content-bearing agents these things wouldn't have content.

Now, Adams and Aizawa claim that there is a broad-consensus in favour of the idea that mentality involves non-derived content, citing the work of Dretske (1981, 1988), Fodor (1987,1990), Searle (1980), and Millikan (1984) as support. But there are arguments to the contrary, such as that presented by Dennett in 'The Myth of Original Intentionality' (1990), and Clark himself doubts that there is a: "clear and distinct sense in which neural representations get to enjoy 'intrinsic contents' of some special kind, quite unlike the kinds of content that figure in external inscriptions." (Clark, forthcoming). Nonetheless, Clark is prepared to accept that non-derived content is the mark of the mental and that all mental states and cognitive processes must possess non-derived content, since he thinks that he can still preserve his arguments for the extension of mental states. I will make the same move here, and explore what follows for EMT if nonderived content is indeed the Mark of the Mental.

We saw in a previous section how Adams and Aizawa sought to challenge EMT by providing a principled difference between internal cognition and putative extended cognition in their examination of Clark

and Chalmers' Tetris cases. To refresh; in case (1) a person attempts to determine the fit of the shapes by way of mental rotation alone, in case (2) the person does so by physically rotating the shapes on a computer screen, and in case (3) the person does so by way of a neural implant. Adams and Aizawa attempt to answer Clark and Chalmers' claims that each case is on a par in so far as cognitive status is concerned by providing a principled difference between the cases. This time they attempt to provide this difference in terms of nonderived content.

> Cognitive processing is, of course, involved in all three cases, but in different ways. (1) and (2) differ in their use of non-derived representations and in the sorts of processes that go on in them, hence (2) does not constitute a 'real world' case of transcranial cognition. (Adams & Aizawa 2001: 54).

This requires a little more fleshing out. Adams and Aizawa state that in case (1) the agent uses mental representations of the blocks to perform the task whereas in case (2) the blocks that are rotated are not representations at all, either derived or non-derived. 'They do not *represent* blocks to be fit together; they *are* the blocks to be fit together.' (Ibid.). Fodor echoes this point stating that: "The world can't be its own best representation because the world doesn't represent anything; least of all itself. The world doesn't mean anything and it isn't about anything; it just is." (Fodor 2009). The bottom line according to these objectors then is that if *cognition* must involve nonderived representation then the manipulation of the on-screen images by pressing a button is not a case of genuine cognition.

The same contrast is said to hold between Inga and Otto with his notebook:

> One obvious difference between the two cases involves non-derived content. Where the symbols written in Otto's notebook have merely derived content, the recollection in Inga's brain has non-derived content. Otto's notes do not, therefore, constitute beliefs or memories. (Adams and Aizawa 2001: 55)

Andy Clark responds to the non-derived content objection in 'Memento's Revenge: The Extended Mind: Extended' (Forthcoming), by posing the following question:

> [M]ust everything that is to count as part of an individual's mental processing be composed solely and exclusively of states of affairs of this latter (intrinsically content-bearing) kind? I see no reason to think that they must. (Clark, forthcoming).

He examines what he believes to be a genuine case of cognition involving the visualizing of Euler Circles. Suppose that we are set some cognitive task that we solve by visualizing a set of Euler circles in order to reach a solution. It is surely the case, Clark argues, that the fact that this set of Euler circles means anything, regardless of whether or not they are visualizations, is a matter of convention. If so, then the content that the visualization of the set of Euler circles has is derived content and it clearly has a role to play in the completion of the cognitive task, it is clearly part of the cognitive process. So his point is that in cases like this we have a case of genuine cognition that involves derived content, which means that perhaps not everything that counts as part of an individual's mental processing must be composed solely and exclusively of states of affairs bearing non-derived content.

Adams and Aizawa meet the challenge laid down by Clark:

> Evidently the problem here is supposed to be that there are some mental

> states that have contents in virtue of a social convention. So, Clark implies that there are *bona fide* cognitive processes that involve derived content… Our view is that Clark's analysis of the Euler circles case is superficial and confused. (Adams & Aizawa, forthcoming).

Adam and Aizawa's problem with Clark's analysis of the Euler circle case is that although social convention is involved in the meaning of the overlap of Euler circles, this is a logically separate matter from what makes an imagistic mental representation of intersecting Euler circles mean what it does. How intersecting Euler circles on paper, for example, get their meaning is a distinct matter from how Euler circles in mental images get their meaning. Intersecting Euler circles may mean set-theoretic overlap by convention but it is by no-one's convention that a particular neural state means intersecting Euler circles.

> It can be a matter of convention that "dog" means dog, that a stop sign means that you should stop, that a person raising a white flag means to surrender, and that a red light flashing means that something is overheating. But, that does nothing to show that it is not the satisfaction of some set of naturalistic conditions on non-derived content that get something in the head to have the meanings of "dog", a stop sign, a white flag, and a warning light. (Ibid.).

This may well be the case but it could still be argued that that there cannot be mental images in which intersecting Euler circles mean set-theoretic overlap unless there were a social convention according to which intersecting Euler circles meant set-theoretic overlap. In a case like this doesn't the meaning of the mental image derive in part from the prior existence of the meaning of external pictures? The meaning of the mental image in this case could be said to depend on the existence of a prior meaning. Adams and Aizawa claim that an argument such as this trades on an ambiguity in the notions of derivation and dependency.

> Insofar as there must be a social convention regarding the intersections of Euler circles in order to have a mental representation regarding the intersections of Euler circles, this is not a fact about the constitution of the content of a mental image of the intersections of Euler circles. It is, if anything, a kind of historical fact. (Ibid.).

So what they are stating is that it is true that without prior social convention involving the intersection of Euler circles meaning set-theoretic overlap one could not have a mental image involving the intersection of Euler circles meaning set-theoretic overlap. But this is just like saying that without the prior existence of cars one could not have a mental image of a car. This, according to Adams and Aizawa is merely a type of historical truth and does not show what a defender of EMT may want it to show; that the content of certain mental items derives, in some relevant manner, from a social convention.

Thus, for the EMT hostile, the image of intersecting Euler circles although dependent on the fact of the prior existence of Euler circles is ultimately grounded in neural traces within the head, which are logically prior to the image. The image of the intersecting Euler circles gets its meaning from a pattern of neural firings within the head; the image must be triggered by neural goings-on with non-derived content, and the understanding must, ultimately, consist in this.

But Clark hones in on a concession made by Adams and Aizawa with regard to the pervasiveness of non-derived content through the whole of the cognitive process:

> Having argued that, in general, there must be non-derived content in cognitive processes, it must be admitted that it is unclear to what extent

> each cognitive state of each cognitive process must involve non-derived content (Adams & Aizawa 2001: 50).

Clark takes this to mean, not unreasonably, that even if there is a requirement for non-derived content in the larger cognitive process, there can be elements of the cognitive process that exhibit derived rather than nonderived content. If cognitive states and processes must involve nonderived content in every stage, then it does, initially at least, seem legitimate to rule out pressing a button to manipulate images on a screen as a cognitive process. Similarly, it would seem legitimate to rule out the notes in Otto's notebook as beliefs and/or memories. But, as Clark points out, the argument is not that pressing a button on a computer screen is a cognitive process, the argument is that is it forms *part* of the cognitive process. Likewise the argument is not that the words in Otto's notebook in isolation count as beliefs and memories, they form *part* of Otto's cognitive state in the case of belief, and part of Otto's cognitive process in the case of memory.

Adams and Aizawa seem to want to consider the external processes in isolation from what occurs within the head but this is exactly what the EM theorist is not claiming. EMT has it that the external components in combination with what happens in the head form part of an extended cognitive system. If you want to apply the nonderived content argument to these examples, then you need to consider the process over the extended system as a whole. Adams and Aizawa do not do this. They consider the external components in isolation from the internal processes and conclude that the external processes do not possess the necessary nonderived content. However, if you examine the extended system as a whole then the nonderived content that Adams and Aizawa require can be found exactly where they expect it to be, inside the head. Thus, the EM theorist argues, Adams and Aizawa cannot reject the examples discussed

above as cases of genuine cognition on the grounds that they don't contain non-derived content unless they consider the external components in isolation, which is a misunderstanding of the EM theorists' position.

However, perhaps Clark was too hasty in using Adams and Aizawa's concession that not every part of every cognitive process need exhibit nonderived content as a licence to include derived content in the larger cognitive economy as a constituent part of a cognitive process that also involves nonderived content. After all if we examine Adams and Aizawa's statement immediately following the concession seized on by Clark we get a better idea of what they have in mind:

> [I]t is epistemically possible that cognitive processes involve representations that include a closed set of non-representational functional elements, such as punctuation marks and parentheses. Such items may be included in the language of thought, based on the manner in which they interact with items having non-derived content. (Ibid.)

So Adams and Aizawa allow that things other than intrinsically content-bearing states of affairs do have a role to play in cognition in the form of certain non-representational functional elements. This is not the same as conceding that derived content can play a role in a larger cognitive process that involves non-derived content. Perhaps Adams and Aizawa mean to rule out derived content from forming any part of a genuine cognitive process while admitting that certain non-content-bearing elements will have a role to play?

In such a case it seems that they can legitimately rule out Otto's use of the notebook as a cognitive process because it will involve derived content in the form of the entries in the notebook. Similarly, if they are correct then they may be able to admit Clark's Euler circle example as a

genuine cognitive process as the meaning of the image is dependent (except historically) on the particular pattern of neural firings that instantiates the image and is not derived from any prior content. We may question where this leaves the Tetris example however. Since Adams and Aizawa and Fodor have argued that the blocks on the computer screen are not representations this means that they cannot be derived representations, which would mean, if this interpretation of the non-derived content requirement is correct, that there is nothing to rule out the manipulation of the blocks on the screen as part of a larger cognitive process that extends from the head into the world.

However, the interpretation of Adams and Aizawa's position whereby they demand that derived content *cannot* form part of a cognitive process is not the correct one it seems. Adams and Aizawa attempt to clarify: "Clearly, we mean that if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive." (Adams & Aizawa, forthcoming).

It seems that Adams and Aizawa are offering non-derived content as a *necessary* condition on a process being cognitive. But a mark of the cognitive that specifies non-derived content only as a necessary condition on a state's being cognitive seems to open the door for the EM theorist with regard to the extension of cognitive processes. If insisting that non-derived content is the mark of the cognitive merely provides a necessary condition for a state or process as a whole to count as cognitive then Clark and Chalmers can argue that there are other aspects (perhaps not representational) of a cognitive process that might extend into the world. This might be the case if they are specially related in some way to the internal non-derived content by fulfilling the additional criteria on an extended putative cognitive resource specified by Clark.

Likewise, a mark of the mental that specifies non-derived content only as a necessary condition for awarding mental status would also seem to open the door for Clark and Chalmers with regard to the extension of mental states into the world; and therefore with regard to the possibility of a genuine EMT. In the next chapter I explore this possibility in returning to examine the fundamental account of mental states that is presupposed by EMT.

Chapter 5

Introduction

In this chapter I re-examine the fundamental account of mental states and their properties that is presupposed by EMT in order to clarify precisely the nature of its claims. Functionalism is taken by many to be the inspiration for EMT, with both sides of the debate accepting that EMT depends on or derives from functionalist principles. I assess, then, the relationship between EMT and functionalism, clarifying that it is analytic, or commonsense, rather than empirical functionalism that Clark relies on to make his case for extended mental states. I also investigate what it is about functionalism that makes it so important for EMT and take a closer look at the nature of mental states according to the functionalist perspective.

If functionalism alone is sufficient to wholly define the propositional attitudes that Clark seeks to extend then EMT is a possibility. However, the propositional attitudes that Clark takes to be the paradigm mental states have two components; the attitude and the proposition or content. It is commonly taken that the attitudinal aspect of mental states with propositional content is functionally defined, i.e. it is functional profile that distinguishes beliefs from hopes and desires, etc. But it is content that individuates a belief *that the Museum of Modern Art is on 53rd Street* from a belief *that water is wet.* And nothing could count as a belief (or any other propositional attitude) unless it possessed content. So unless one buys the idea that content reduces to the functional, which is not a credible proposal, then functionalism tells us something about what is necessary but not *sufficient* for being a mental state such as a belief *that the Museum of Modern Art is on 53rd Street*. If pure functionalism (i.e. a complete account of the properties of propositional mental states in

functionalist terms) were true then syntactic minds would be a real possibility. But if we rule out purely syntactic minds because they do not carry the mark of the mental then we rule out a purely functional criterion of mind. So, at best, the functional aspects of propositional attitudes *might* extend, but if the content doesn't then the defining properties of mental states do not extend.

Thus, if non-derived content is the real mark of the mental and it is internal, then argument for a genuinely interesting EMT is going to be difficult. In presenting his argument for EMT as based on functionalism Clark focuses exclusively on extending the formal properties that fulfil the functional role of mental states, ignoring this mark of the mental. Despite his claims that he wants his EMT to be viewed as an "environmentally extended case of narrow content" (Clark, forthcoming), Clark contends he can allow that non-derived content is the mark of the mental and that it is internal and still argue for the extension of some mental states into the world. He does this by invoking the special role that external factors (like Otto's notebook) sometimes play in the generation of action. But this move only supports a case for the extension of the machinery of mind in virtue of the role that external factors play in enabling propositional attitudes and it is difficult to see how this EMMT could definitively secure a full-blown EMT.

In response to these criticisms Clark could retreat from the claim regarding the extension of mental states (since he doesn't argue for the extension of the defining properties of mental states) to the extension of cognitive processes that, perhaps, need not themselves display the mark of the mental. But in such a case there will be no reason to think that EMT concerns an extension of the mental in any interesting sense. Once more we are faced with the conclusion that EMT is not as advertised and reduces to EMMT. This sets the stage for the next chapter where I will,

by way of conclusion, review some new movements in philosophy of mind that open up the possibility of taking a more radical line than Clark suggests with regard to extending mental states, in this way exploring what it would take for there to be a more interesting version of EMT.

Functionalism and EMT

Functionalism in the philosophy of mind is typically taken to be an internalist position—although, as I demonstrate, this is not logically required. Wilson, for example, points out that functionalism has had a large role to play in making internalism the default view of the nature of mind:

> With the rise of functionalism, the claim that mental states are realized in physical states of the brain became part of the received wisdom in the mind-body relationship. (2004: 101)

He argues that the two most prevalent ways of understanding functionalism within cognitive science—computationally and in terms of analytic decomposition—are especially amenable to internalist readings (or individualistic readings as Wilson puts it):

> Computational processes, conceived as operating solely on the syntactic properties of mental states, have been plausibly thought to be individualistic and it is natural to view analytical decomposition as beginning with a psychological capacity, such as memory or depth perception, and seeking the intrinsic properties of the organism that create and constitute that capacity. (Ibid: 95-96)

Similarly, Jackson and Pettit point out that "it is alleged that functionalism is wedded to the inner picture" whereby "mental processes are essentially inner processes" (1988: 382). They cite McDowell as

supporting such a view and indeed he suggests that functionalism might be the modern day replacement for Cartesian immaterialism in fulfilling the demand for the autonomy of the mind from the world (McDowell 1986). Rowlands (2003b) likewise acknowledges that functionalism is usually understood as a form of internalism but points out that this need not be the case.

He argues that functionalism is only wedded to internalism if we specify the functional roles that characterise mental states narrowly; as beginning and ending at the boundary of the subject (i.e. the skin). Functionalism defines mental states as the states that occupy certain causal roles between input, output and other similar states. But, as Jackson and Pettit (1988) argue, it is possible to specify inputs and outputs as happenings that are outside the skin and, in fact, they recommend adopting just such a strategy in order to deal with the existence of broad content. Note that if functionalism alone were sufficient to account for the nature of mental states then this broad functionalism of Jackson and Pettit would be sufficient to secure the extension of mental states into the world.

Now, what matters, from the functionalist perspective, is simply that the functional role is fulfilled, not how it is filled or what fills it. In other words all that matters is that the job gets done and *what* gets it done is incidental. This implies that functional states are multiply realizable and also that functionalism should be locationally neutral with respect to what fills the functional role. One may adopt the position that what in fact fills the functional roles characteristic of mental states are internal brain states, but there is nothing within the pure functionalist perspective that requires such a move.

In any case, since the publication of 'The Extended Mind' in 1998 many have recognised that functionalism is not necessarily an internalist

position and the links between functionalism and the possible extension of mental states and cognitive processes into the world have been the topic of much discussion, with Sprevak even going so far as to claim that "functionalism entails that cognitive processes *do* extend in the actual world" (2009: 503). He argues for this claim by pointing out that a functionalism that specifies the functional identity of mental states and cognitive processes at a course enough level of grain to preserve the 'Martian intuition' must include counting on your fingers as a cognitive process and "Otto's notebook…as an extended belief" (Ibid: 510). The Martian intuition is the intuition that, for any given *type* of human mental state, having a human physical and fine-grained psychological makeup is not necessary in order to have that state, thus allowing Martians of radically different physiology to be in type-identical mental states to us. Sprevak goes on to recommend the rejection of functionalism on the basis that it entails just this sort of cognitive and mental extension but in this he is unique among those currently writing on EMT and, as we will see, most hold that functionalism at least opens up the possibility of EMT (although we must be careful to remember that it will only do so *if* functionalism can tell the complete story about mental states).

In an interesting treatment of EMT Shapiro (2007) correctly points out that the possibility of extending mental states and cognitive processes into the world must rest on a prior theory of mind. So, for example, if one endorses a mind/brain identity theory then the possibility of extending mental states beyond the limits of the brain is a non-starter. EMT must rest on a theory of mind that does not rule out its very possibility. Shapiro goes on to suggest that the dispute over the possibility of EMT ought to rest on a shared agreement about which theory of mind is correct:

> The controversy is interesting only insofar as its participants share a view about what minds are but disagree over how to draw the mind's

> boundaries. Fortunately, many involved in the dispute seem committed to a common theory of mind, viz. functionalism. (Ibid: 5)

Shapiro's contention that both sides of the debate share intuitions regarding a theory of mind is borne out by what is in the literature. Chalmers, for instance, denies that EMT requires that *all* mental states be defined (even in part) by the causal roles they play (Clark 2008b: xv), but does acknowledge that some measure of adherence to a functionalist theory of mind is required if the possibility of EMT is to be acceptable:

> All one needs is the very weak functionalism captured in the Parity Principle: roughly, if a state plays the same causal role in the cognitive network as a mental state, then there is a presumption of mentality.[31] (Ibid.)

Chalmers' reluctance to accept an all embracing functionalism about the mind comes from his intuition that functionalism about consciousness is implausible. This is why you will find no argument for the extension of consciousness in Clark and Chalmers' paper: "[I]t seems far from plausible that consciousness extends outside the head in these cases" (1998: 10).[32] But although Chalmers may recoil from endorsing a fully fledged functionalism, he does recognise that in order for mental states to extend in the manner argued for by EM theorists, functionalism must be true of at least some mental states – i.e. propositional attitudes. This seems to be the general consensus among those writing on the topic and

[31] Again, we must be careful to remember that this will only be the case if functionalism tells the whole story about mental states. For example, anyone who rejects the possibility of a purely syntactic theory of belief will reject the idea that there is a presumption of mentality in such cases. I explore this in more detail in a later section.
[32] Clark (2008) in contrast has since professed that he is tempted by a functionalist account of conscious mental states. But there are good reasons for thinking that qualia, as definitive of conscious mental states, will not functionalise and the same may be true of content, as definitive of propositional attitudes. I explore this in more detail in a later section. In 'Spreading the Joy?' (2009) Clark argues against the extension of consciousness.

this is where EMT is thought to gain its hold. Clark, certainly, is firmly ensconced in the functionalist camp, often referring to himself as a 'good old-fashioned functionalist.'

In response to a charge by Adams and Aizawa (2007) of being a closet behaviourist Clark points out that his version of EMT is: "not behaviourism but (extended) common sense functionalism" (2008b: 96). And when highlighting the crucial differences between embodied and extended accounts of cognition he argues that:

> It is the tension between an extended, situated-reason (sic) friendly version of good old fashioned functionalism, and something more fundamentally fleshy: the idea that features of the body make a special and in some sense non-negotiable contribution to mind and mentality. (Ibid: 51)

The tension referred to is between what Clark calls bio-chauvinism (expressed in this case by the thesis of the embodied mind, which accords a special role to the body in realizing mental states) and the functionalist indifference of EMT to the physical realizers of mental states. It is in this indifference that the appeal of functionalism for the EM theorist lies. According to Clark this indifference of functionalism to the nature of the realizer should include an indifference to the location of the realizer, i.e. functionalism should be locationally neutral, and this is the key feature of functionalism so far as EMT is concerned.

As we saw in chapter three Hurley (1998b) argues that whether or not a state or processes is 'central' and worthy of being called a mental state or process should be specified functionally rather than by the location of the process. This is what, she claims, allows her to postulate the extension of vehicles into the world. And although Rowlands does not explicitly

invoke functionalism in presenting his version of EMT, he does rely on the basic functionalism that is captured by the parity principle in order to make his argument work. As we saw in chapter 2, Rowlands argues that part of the requirement for a process being a cognitive process is that it involves the manipulation of information-bearing structures and that the process is essential to the completion of the cognitive task. It is irrelevant, on this view, where the process takes place or what the realizer of the process is, what matters is that the process fulfils the functional role that contributes to the completion of the cognitive task in question.

Adams and Aizawa also recognise the reliance that the arguments of the EM theorist have on functionalism and although they reject EMT as an empirical fact they endorse functionalism and consequently recognise that the characterisation of cognitive processes in terms of functional role leads to the extension of the realizers of those roles as a logical possibility:

> The multiple instantiability of functional categories is a familiar element in contemporary cognitive science. Applying this to our rather orthodox conception of the mark of the cognitive, one might think that non-derived representations and the sorts of functional processes that are found in the brain might also be instantiated in systems that cross the boundaries of the brain. One might think, therefore, that for all the proposed conditions on the mark of the cognitive show, transcranial cognition is a live possibility. We agree with this line. Transcranial cognition is a live possibility. Our view is simply that, as a matter or boring contingent empirical fact, transcranial and extracranial cognition are not commonplace. (Adams & Aizawa 2001: 60)

Adams and Aizawa's mark of the *cognitive* is, as we saw in the previous chapter, non-derived content. However, although it is clear that for propositional attitudes content is the defining feature, it is possible to

specify cognitive processes non-intentionally such that it is at least debatable as to whether or not non-derived content is also the mark of the cognitive. If so then a pure functionalist account of cognitive processes may be possible, in which case it may be correct to say that transcranial cognition is a live possibility. However, if non-derived content is the mark of the mental it is difficult to see how the same could be true for mental states since functionalism will only give us a partial account.

Daniel Weiskopf takes a similar line to Adams and Aizawa and highlights the derivation of the arguments for EMT from functionalism, pointing out that the arguments of Clark and Chalmers in particular are clearly dependent on the pure functionalist principle that what makes something a mental state is the functional role that it plays, regardless of its location. Weiskopf points out that this principle is: "no more than the elementary functionalist tenet that to be a belief is just to play the causal/functional role of belief." (2008: 266) but it should be emphasised that there are not many who would endorse this pure functionalist account of propositional attitudes, recognising that the essential contentful aspect of propositional attitudes is not amenable to functionalist explanation. Nonetheless, Weiskopf argues against the internalist tradition, which has it that the realizers of the causal/functional role of a belief (or indeed of any mental state) are internal, that the possibility of more exotic realizers of mental states is a live one:

> Unusual realizers are a staple of the functionalist literature. The hybrids described by advocates of extended minds differ only in lying outside of the normal brain-body system… [I]n itself it is simply an instance of an unusual realization of a mental state, and thus a fixture familiar to functionalists. So there is a sense in which the extended mind thesis should not be seen as especially radical. Functionalists have all along

> been committed to the possibility of extrabodily states playing the role of beliefs and desires.[33] (Ibid: 266-267)

The links between functionalism and EMT are well established in the literature then. But there is more than one version of functionalism and we can question which version of functionalism is most amenable to the purposes of the EM theorist. As we saw above, Clark advertises that his EMT is extended *commonsense* or analytic functionalism.

Commonsense functionalism must be distinguished from empirical (or psycho) functionalism, a distinction that Clark (2008b) credits to Braddon-Mitchell and Jackson (2007). On the commonsense functionalist view, we attribute a particular mental state, such as the belief *that the museum of Modern Art is on 53rd Street*, based on common, or folk, knowledge about mental states; on what causes actions, given certain perceptual inputs and assumptions concerning certain other of the individual's mental states. For example, Otto hears about an exhibition at the Museum of Modern Art (the perceptual input), forms a desire to see it (mental state) which relates to other mental states, i.e. beliefs, hopes, etc., such that he forms the intention to go to the exhibition and this causes him to set off in the direction of 53rd street (action). On the commonsense view we would not hesitate under these conditions to attribute the belief *that the Museum of Modern Art is on 53rd Street* to Otto. This commonsense functionalism has it, then, that to believe *that the Museum of Modern Art is on 53rd Street* is to be in a state of mind that plays the belief-role in this complex network of states. Since playing this sort of role in the theoretical network is what conceptually defines 'belief' this view is also referred to as analytic functionalism. Analytic functionalism looks to the commonsense mental state ascriptions of the folk and says

[33] Once more, this last sentence is true only for someone who endorses a pure functionalist account of mental states and there are not many (any?) who do.

that mental states are those entities that fill the functional roles characteristic of these explanations. This is, as Clark (2008b) points out, a rather course-grained functionalism about mental states—we can and do attribute mental states in a way that abstracts away from the implementational details of what realizes them and being pitched at this level of abstraction is what makes commonsense functionalism ideal for the purposes of the EM theorist.

In contrast, empirical functionalism claims that mental states are: "just those entities with just those properties, postulated by the best *scientific* explanation of human behavior" (Levin 2009). This empirical functionalism *may* also make use of our commonsense folk-psychological mental state ascriptions but not as providing a full course-grained account of what is required to fill the particular functional role of a mental state. Instead, empirical functionalism *might* use commonsense mental state ascriptions merely as a starting point for a more fine-grained analysis of the functionally defined mental state. Thus, while one could use commonsense ascriptions initially to pick out potential candidates for scientific investigation the aim of this sort of functionalism is to discover, scientifically, what the nature of putative mental states really are. This is not thought to be fixed by a commonsense theory; it is not folk psychology that defines our understanding mental state concepts on this view. So, even if common sense might initially enable us to pick out mental states by targeting things that play such and such a role (i.e. the role of pain) our conceptual understanding of such mental states is determined *wholly* by the outcome of empirical investigation. Scientific enquiry will discover what pains are and exactly how they function in the mental economy of particular species according to the empirical functionalism. Thus, it will be an empirical matter to discover just what the role of pain is in humans and what plays that role (e.g. it may be that to be in pain to have one's C-fibres firing in such and such a way, for

example). It is the business of science to discover exactly what role pain plays and what exactly happens to play that role on this view.

Thus, one can advocate an empirical functionalism that does not take the common sense mental state attributions of the folk very seriously at all. Such a view allows that we may discover things about mentality that are quite radically at odds with our everyday intuitions about the properties of the mind and opens the possibility that our commonsense mental state ascriptions do not in any way accurately describe the ontology of the mind, e.g. there may be no such things as beliefs and desires (cf. Churchland 1981).

Because, *prima facie*, empirical functionalism takes a more fine-grained stance on what it is for two states to be functionally similar, than the more course-grained approach of analytic functionalism that abstracts away from the implementational details, it has been argued that the methodology of empirical functionalism leaves it open to the charge of chauvinism (i.e. it would not licence what Sprevak calls the Martian intuition, cf. Block 1980). Empirical functionalism may not be as liberal in its attributions of mentality as analytic functionalism, which would seem to make analytic functionalism a better candidate for the purposes of the EM theorists. But there is also an additional worry here that the empirical functionalist theories of our psychology may move away from the belief/desire explanation of action that characterises the mental attributions of the folk such that "it will be hard to take psycho-functional theories as providing an account of our mental states, rather than merely changing the subject" (Levin 2009). Given that Clark is seeking an account that can extend propositional attitudes into the world it seems that an empirical functionalist account of mind, which may not even licence the commonsense mental state attributions of the folk, is not the way to go. This is not to say that a pure empirical functionalist account of

mind could not licence the extension of realizers into the world, rather the point is that we may find that the best scientific explanation of our psychology does not include such things as the beliefs and desires that Clark seeks to extend.

The appeal to Clark, then, of commonsense functionalism is first of all that it allows him to attribute the belief *that the Museum of Modern Art is located on 53rd street* to Otto. If what it means to believe *that the Museum of Modern Art is located on 53rd street* is simply to go towards 53rd street after hearing of a new exhibition at the Museum of Modern Art given a desire to see this exhibition, then, on this view, Otto has this belief. On this view the fact that Otto fulfils the necessary commonsense functional role is all that is required to attribute the belief to Otto:

> It is the course or common-sense functional role that, on this model (unlike that of empirical functionalism), displays what is essential to the mental state in question. (Clark 2008b: 89, emphasis mine).

But if the EM theorist is to argue not only that Otto believes *that the Museum of Modern Art is on 53rd Street* but also that Otto's belief extends into the world, then he must enquire as to what in reality realizes this functional role.

> [W]e may go on to seek a much more fine-grained description of the actual flow of processing and representation in the (possibly extended) physical array that *realizes* the course functional role itself. (Ibid: 88-89)

It is a further question, then, as to what realizes this course functional role; a question that is to be answered by empirical investigation. Clark's argument is that empirical investigation will reveal that external

information-bearing structures such as Otto's notebook are often essential components of the realizer. If analytic functionalism is correct and mental states are multiply realizable then there is no reason to restrict empirical investigations to locations within the brain or body. Instead, we should also examine the external environment in order to ascertain if some external feature makes an essential contribution to realizing this particular mental state. If this turns out to be the case then functionalism demands that we accord it 'fair treatment' (Sprevak 2009) and designate it as part of the realizer of that mental state. If some internal feature played the role we would not hesitate to label it as part of the realizer of a mental state. This is the message of the Parity Principle.

However, it is not clear that the extension of *part* of the realizer of a mental state will be sufficient to secure the extension of the mental state into the world. The claim here is not that the complete realizer must be externally located in order for mental states to extend (this would be the External rather than Extended Mind), rather the claim is that if the realizer of the functional role of a belief can account only for the attitudinal aspect and not the content and if the content remains internal (as Clark concedes) then in what sense do we really have a case for the extension of a mental state into the world as opposed to a further case for the extension of processing into the world that Clark wants to move beyond?

The question for Clark, then, is in what sense does he think that the commonsense role displays what is *essential* to the mental state? If he means that fulfilling the functional role characteristic of the mental state is *necessary* for the attribution of the mental state then he is not likely to meet much opposition with regard to this claim but nor will he secure the extension of mental states into the world. If, on the other hand, he means that fulfilling the functional role characteristic of the mental state is

sufficient for the attribution of the mental state then he is no longer talking about the extension of propositional attitudes and will meet substantial resistance. In the next section I look in more detail at the typical functionalist account of mental states to demonstrate why this should be the case.

Functionalism and Mental States

A good way of getting at an account of how functionalism characterises mental states is to approach via its historical antecedent; behaviourism. According to behaviourism when we say that a person is in a particular mental state what we mean, roughly, is that the person acts or is disposed to act in a certain way given a certain stimulus. In this way behaviourism offers a basic stimulus-response model to explain human action; defining mentality without 'the inner' thus denying, in the process, the reality and so the causal efficacy of mental states (Lewis 1966). On this view the actions of an individual can be explained entirely by reference to various inputs.

Behaviourism abandoned the idea that mentality is essentially linked to the first-person perspective, which had dominated philosophy of mind up to that point. Private, inner states of the sort postulated by dualism are, according to the behaviourist, anathema to the project of making psychology into a respectable science. Because non-physical mental states are not observable they are not measureable and so should have no place in a scientific psychology. So, on one behaviourist account what it means to be in a certain mental state is to behave or act in a certain way (or to be disposed to) in response to certain stimuli. Such responding could be highly complex and multi-tracked. Thus one way to be in pain is to jump or scream or wince (or all three of the above) when poked with a hot needle, say. On this account, understanding what it is to be in pain

would require an understanding of all the various ways one might react to various stimuli. On another, slightly different, behaviourist account; that a person is in pain is made contingently true or false by how that person responds or is disposed to respond given the 'poked with a hot needle' stimulus. In either case the behaviourist does not require the postulation of internal mental states to explain what it is to be in pain, or what it is to believe *that p*.

Behaviourism is criticised for ignoring the 'intractable residue' of consciousness by failing to do justice to the reality of the inner. But, more significantly perhaps, behaviourism received widespread criticism because a complete behavioural analysis of certain mental state ascriptions cannot be given without appeal to some other mental state ascription. So, for example, a belief *that the Museum of Modern Art is on 53rd street* cannot properly be accounted for in behaviourist terms because without some other mental state, such as a desire *to see the exhibition currently on display at the Museum of Modern Art*, there will be no action. Beliefs, on their own, are not sufficient to cause action; there must be some accompanying desire. The behaviourist's attempts to characterise mental state ascriptions by reference to stimuli and behaviour alone, without reference to other mental states is, then: "either circular or radically incomplete as analyses of the mental generally" (Lycan & Prinz 2008: 4).

Functionalism was proposed in part as a solution to this dilemma of trying to balance the scientific requirement for verifiability with the need to appeal to mental states in order to justify certain mental state ascriptions. Instead of focusing solely on the inputs and the outputs—the stimuli and the responses—analytic functionalism postulated that the folk posit the existence of a set of theoretical states between the inputs and the outputs. This theory assumes the reality and causal efficacy of mental

states and that these are investigable by science. In this way functionalism sought to tackle the problem of making psychology into a respectable scientific endeavour, not by denying the reality of mental states but by viewing commonsense mental state attribution as the postulation of theoretical entities that come between the input of perception and the output of behaviour. Inner mental states provide the meat in the sandwich between the stimulus and the response.

It is commonly taken that there are two main sources of inspiration for this functionalist characterisation of inner mental states. The first of these is computer science where the distinction between the hardware and software was seen as a potentially useful aid for understanding the relationship between mind and body and the second is a view about the nature of theoretical terms as expressed by David Lewis, which he applied to commonsense mental state ascription. It is this latter inspiration for functionalism that I focus on since it is commonsense, analytic functionalism that Clark espouses.

Lewis held that theoretical terms were defined implicitly by the theories in which they were introduced because they are definable by reference to the causal roles they occupy within the theory (1972). Lewis invites us to think of commonsense mental state ascription in just this way. On this view mental state predicates like belief and desire are theoretical terms that have been introduced within this folk-scientific theory in order to explain the actions of individuals. And, like theoretical terms in general, the theoretical terms of this folk theory are definable, says Lewis, in functional terms by reference to their causal roles. So, folk psychology associates each mental state with a typical causal role and in this way folk psychological explanations concern the causal efficacy of mental states by stating, with reference to its relations with various inputs and other mental states, what behaviour the state in question is likely to cause.

Because the folk engage in this practice of explaining actions by means of appeal to mental predicates such as beliefs and desires, they are said to be committed in some sense to the existence of beliefs and desires as theoretical entities.

It may seem as if the attempt to explain mental states functionally in this way is circular, since the explanation of any mental state requires reference to at least one other mental state and so presupposes other mental states. But Lewis' idea is that we can simultaneously define all mental states at once in non-mental terms by ultimately substituting the theoretical, or T-terms (i.e. the mental state terms like beliefs, desires etc.), with terms of another, non-mental, vocabulary.

In order to do this Lewis advises us to: "Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses" (1972: 226), and combine them into one long sentence that Lewis calls the 'postulate' of our folk psychological theory. We can give an example of what Lewis has in mind with the Otto case that we have been looking at. The postulate of the belief/desire pair regarding the Museum of Modern Art for Otto might look something like this:

> When Otto has a desire to see the Dali exhibition currently on display at the Museum of Modern Art and Otto has a belief that the Museum of Modern Art is on 53rd street he will walk towards 53rd street and when Otto has a desire to see the Dali exhibition currently on display at the Museum of Modern Art and Otto does not have a belief that the Museum of Modern Art is on 53rd street he will not walk towards 53rd street (*ceteris paribus*).[34]

[34] This sentence does not properly characterise this particular belief/desire pair in its entirety but it is sufficient to illustrate the idea.

We can see immediately that the sentence consists of several of what Lewis would call T-terms, like the desire *to see the Dali exhibition* and the belief *that the Museum of Modern Art is on 53rd street*. It also consists of other terms that Lewis calls O-terms. An O-term is any term other than a T-term, a term that was already in use and understood before the advent of the new theory T and its T-terms. The idea is that we can replace the T-terms in the postulate with variables and hold onto the O-terms enabling us, eventually, to describe the T-terms in language that does not make use of mental state terms.

So, we can re-write the postulate of our theory as follows with the T-terms replaced by variables in the following manner:

> When Otto has x_1 and has x_2 he will walk towards 53rd street and when Otto has x_1 and does not have x_2 he will not walk towards 53rd street.

What we have here is a two-place predicate that we can call our theory, T, and the T-terms of T are the desire *to see the Dali exhibition* and the belief *that the Museum of Modern Art is on 53rd street*, represented by the variables x_1 and x_2 respectively. If the first sentence above is our theory, T, then we can re-write the postulate of our theory as: T(desire to see the Dali exhibition, belief that the Museum of Modern Art is on 53rd street), or with the variables in place of the T-terms as: $T(x_1,x_2)$. We can now write the Ramsey sentence for our theory T, which would look something like the following:

$$(\exists x_1)(\exists x_2)\ T(x_1,x_2)$$

And which says that there exist two things, call them x_1 and x_2 respectively, such that if Otto has x_1 and Otto has x_2 and Otto hears of a Dali exhibition at the Museum of Modern Art he will walk towards 53rd

street etc. This demonstrates, says Lewis, how our theory, T, is committed to the existence of entities like beliefs *that the Museum of Modern Art is on 53rd street* and desires *to see a Dali exhibition*. Lewis explores the modified Ramsey sentence and modified Carnap sentence of folk psychological theory but that need not concern us here for the point is made; on the commonsense or analytic functionalist view, attributing mental states as explanations of action commits one to the existence of entities such as beliefs and desires.

Lewis argues that our complete folk psychological theory can be given the same treatment as our theory regarding Otto's belief/desire pair and so the T-terms of our folk psychological theory—the various beliefs, desires, feelings, and experiences—can be functionally defined in terms of their causal roles and the variables (standing in for the mental state terms) can eventually be replaced by terms from some other vocabulary, possibly those of neuroscience. Thus, what we have been looking at here is really only stage one of the project for Lewis.

> [I]t is possible in principle to replace the free variables (standing for mental state terms) with terms in some other vocabulary standing for non-mental phenomena. Thus the stage one work of collecting and analyzing folk platitudes is a prelude to determining what, if anything, in the natural world (e.g. as identified by neuroscience or physics) might play the sorts of network roles of the mental states identified by the folk theory. (Hutto, forthcoming)

So, according to Lewis if we define a particular mental state in terms of its causal role and we can identify a neural state as fulfilling or realizing this causal role then a particular neural state will be token identical with a particular mental state in some instances:

> Mental state M = the occupant of the causal role R (by definition of M).

> Neural state N = the occupant of the causal role R (by the physiological theory).
> ∴ Mental state M = neural state N (by transitivity of =). (Lewis 1972: 249)

Though we may identify the token mental state with the neural state that happens to occupy the causal role R on a particular occasion we must be careful not to endorse the sort of bio-chauvinism that Clark argues commonsense functionalism is supposed to avoid. Although Lewis argues that any causal role R identified by the definition of any Mental state M may as a matter of fact be occupied by a particular neural state, he is not committed to identifying mental states with neural states, and acknowledges the multiple realizability of mental states:

> In general, or in the case of a given species, or in the case of a given person, it might turn out that the causal roles definitive of mental states are occupied by different neural (or other) states in different organisms. (Ibid: 285)

Thus, functionalism affords the opportunity to token identify the mental state with whatever happens on some occasion to realize the causal role definitive of the mental state. However, to emphasise the point, what it is to be in a particular mental state is *solely* given by the causal role on Lewis' account:

> [T]he definitive characteristic of any experience as such is its causal role. The definitive causal role of an experience is expressible by a finite set of conditions that specify its typical effects under various circumstances. By analytic necessity these conditions are jointly true of the experience and jointly distinctive of it. (Lewis 1966: 19-20)

In this way Lewis sees causal role as capturing what is essential to be a certain kind of mental state. So long as some physical realization state can fulfil the conditions for occupying this causal role then it is a mental state and whatever it is that occupies this causal role in a given instance is to be identified with the mental state. Clark and Chalmers appear to endorse both of these sentiments. Remember Chalmers' statement that: "if a state plays *the same causal role* in the cognitive network as a mental state, then there is a presumption of mentality" (Clark 2008b: xv, emphasis mine) and Clark's statement that: "It is the course or common-sense functional role that, on this model displays what is *essential to the mental state in question*. (Ibid: 89, emphasis mine)

The question here, however, that I flagged in the previous section is what exactly Clark (and Lewis) means by 'essential'. If he means that a complete account of propositional attitudes can be given solely in terms of causal role then he is ignoring content, which most would agree is the more important story that needs to be told with regard to propositional mental states. But this appears to be the strategy since, for Clark, what determines the attribution of a particular mental state to Otto, or anyone else for that matter, is that the functional role definitive of that mental state is occupied. This licenses, says Clark, the ascription of the same belief, the belief *that the Museum of Modern Art is on 53rd street* to both Otto and Inga, since the ascription of this belief is said to depend on the role that the information plays for each of them, independently of however this information might be realized.

Clark and Chalmers also appear to endorse Lewis' position on token identifying the mental state with whatever realizes the functional role on a particular occasion since they argue that should the realizer of the belief ascribed to Otto extend beyond the boundaries of the organism then the mental state ascribed to Otto also extends. This strategy is nicely captured

by Rupert, who gives what he thinks might be a typical argument for EMT that is based on functionalism:

> P1: A mental state of kind F is realized by whatever physical state plays the functional role that is characteristic (or metaphysically individuative) of F.
> P2: Some realizations of functional mental state kinds have physical components external to the organism.
> P3: A mental state extends to or includes all components of its realization.
> C: Therefore, some mental states extend beyond the boundaries of the organism. (2004: 34)

This strategy is a sound one provided that P3 is true. However, as I have flagged in the course of this chapter and as we will see in more detail in the next section, there are good reasons to think that this is not the case since there are questions as to whether or not the functionalist account of mental states captures all that is definitive of the mental. In particular, if non-derived content is the mark of the mental (and, as we saw in the previous chapter, Clark seems willing to concede) then in what sense can the truly mental be said to extend if non-derived content is internal? The EM theorist also faces a related problem in that distinguishing what is constitutive of the realizer of a mental state from what merely enables the realizer by making an essential causal contribution may be problematic. It is to this problem that I turn next.

EMT and Mental States

The EMT account of extended mental states as explicated by Clark and Chalmers has it that the location of the realizer of the functional role that is characteristic of a particular mental state determines the location of that mental state. But anyone wishing to make such a claim faces a problem in

identifying the extent of the realizer of the functional role. Although Shapiro points out that arguments about EMT are interesting only insofar as both sides of the debate presuppose functionalism, he argues ultimately that "functionalism is the *wrong* perspective from which to judge the merits of the extended cognition program" (2007: 6). The reason Shapiro thinks that this is the case helps to highlight a possible explanation as to why the EMT debate has begun to stagnate recently: functionalism makes arguments for or against EMT too easy. This is because functionalism is "ill-equipped to answer a boundary problem that confronts decisions about the extent of a property's realization" (Ibid.).

The boundary problem as described by Shapiro is that functional descriptions do not facilitate a distinction between the realizer of a particular functional role and mere causal contributors to, or enablers of, the realizer of a particular functional role. If Shapiro is correct in this then even if the extension of the realizer of the functional role definitive of a mental state were sufficient to ensure the extension of that mental state there would be no fact of the matter, from the functionalist perspective alone, that could identify the extent of the realizer.

> On reflection, it is not surprising that functionalism is ill-equipped to resolve the questions that the boundary problem raises. Functionalism prescribes a way to individuate kinds or properties. However, the functional roles that define functional kinds or properties do not provide direction for distinguishing those parts of a system that realize a particular functional role from those parts of a system that causally contribute to the realizer's capacity to fill a particular functional role. (Ibid: 11)

We can better illustrate Shapiro's point with an example. We can distinguish the heart, as a functionally defined entity, from the heart *qua*

structural organ located in the chest—what you give/receive when you are involved in a heart transplant. Shapiro labels hearts as functionally defined entities; $hearts_f$ and hearts as structural organs; $hearts_s$ – arguing that with regard to $hearts_f$ we can identify several different and equally legitimate potential realizers.

The functional role that $heart_f$ defines is, according to Shapiro, something like: *that which pumps blood through the body*. Thus, one possible candidate, and perhaps the most obvious one, is a token of $hearts_s$. But although a $heart_s$ may be the most obvious realizer for a $heart_f$ it is not the only potential realizer. A $heart_s$ cannot perform its function without arteries, veins, and capillaries that carry the blood around the body. Why not specify the realizer of $heart_f$ as a token of $heart_s$ plus these blood vessels since they have an essential role to play in the fulfilling of the functional role definitive of a heart? Clark and Chalmers argue, in much the same way, that Otto's notebook has an essential role to play in fulfilling the functional role definitive of the belief *that the Museum of Modern Art is on 53rd Street*.

Alternatively, one could also move in the opposite direction with respect to the specification of the extent of the realizer of a $heart_f$. Oxygenated blood from the lungs enters the $heart_s$ at the left atrium and is subsequently pumped through the body from the left ventricle. De-oxygenated blood that has already been circulated around the body re-enters the heart at the right atrium and is pumped to the lungs for re-oxygenation by the right ventricle. Thus, if the functional role of a $heart_f$ is to pump blood through the body one could specify the left ventricle as the realization of a $heart_f$ with the other three chambers of the heart providing a necessary causal contribution, or enabling role, with regard to the left ventricle fulfilling its function.

From the functionalist perspective alone, each realization of $heart_f$ seems equally valid. Further considerations, from medical science or our evolutionary history for example, may ultimately resolve the issue but there is nothing within the functional definition of a heart that will settle the matter. Shapiro argues that the same holds with regard to functional definitions of mental states.

> Clark and Chalmers' assurance that a sentence in Otto's notebook is a realization of one of his memories, rather than a causal contributor to something in Otto's brain that might be a better candidate for the memory's realizer, at least requires further support. For present purposes, the fact that functionalism does not resolve the realization question is enough. (Ibid: 12)

Clark and Chalmers could argue that because it is Otto's *dispositional* belief that extends in virtue of the location of the relevant sentence in his notebook means that the matter may not be as straightforward as Shapiro argues. They point out that in order for us to attribute the dispositional belief *that the Museum of Modern Art is on 53rd Street* to Inga it is not required that she has that belief as an occurrent mental state. All that is required for us to attribute this dispositional belief to her is that she would be disposed to act in an appropriate manner with respect to 53rd Street given a desire to go the Museum of Modern Art—that the belief would become occurrent in the appropriate circumstance, if you will. On this account a dispositional belief is viewed as something like a static piece of encoded information that is poised in storage to guide action as and when required. Clark and Chalmers would argue that the sentence in Otto's notebook plays exactly this role for Otto and if we attribute the dispositional belief *that the Museum of Modern Art is on 53rd Street* to Inga, then we must also be willing to attribute the same dispositional belief to Otto.

However, it is far from clear that this argument will be sufficient to secure the extension of Otto's belief. To see why consider that for Otto, the proximal cause of his action is the internal representation of the sentence that he reads in his notebook, not the sentence in the notebook itself. If Otto doesn't read the sentence and he doesn't understand it then he won't act. His access to the notebook might very well be, as Clark and Chalmers argue, automatic and subpersonal in just the same way as Inga's access to her memories is but this does not mean that Otto doesn't represent the contents of the notebook internally. Thus, for Otto, the proximal cause of his action is an internal state, and we have no reason to attribute the dispositional belief *that the Museum of Modern Art is on 53rd Street* to Otto even though we may here reasonably attribute the occurrent belief to Otto *that the Museum of Modern Art is located on 53rd street* after he has consulted the notebook.

One principled difference between the Otto and the Inga cases, then, seems to be that for Inga the passive, stored information on the basis of which we attribute the dispositional belief *that the Museum of Modern Art is on 53rd street* to her is poised to be the proximal cause of her behaviour. For Otto this is not the case; the sentence in Otto's notebook gives rise to an internal state, a representation of the sentence in his notebook, which is the proximal cause of his action. Thus, we have no reason to regard the sentence in Otto's notebook as a constitutive part of the realizer of Otto's belief as distinct from the sentence making an essential causal contribution to the realization of Otto's belief, which is internal.

We also have no reason, then, to regard Clark and Chalmers' argument as being for EMT as distinct from EMMT. The problem, from the EM theorists' perspective, as pointed out by Shapiro, is that functionalism alone cannot provide the means by which to rule in favour of either of the

two positions. Thus making arguments for or against EMT as based on functionalism alone, too easy. We have been given no good reason, other than the indifference of functionalism to both the nature and extent of the realizers of mental states, to regard Otto's belief as extended. But, as we have discovered, we have at least equally good reason to deny the dispositional belief *that the Museum of Modern Art is on 53rd street to Otto* and to limit the extent of the realizer of his occurrent belief on just these same grounds. As we will see there are further points to tell in favour of just such a move, but for present purposes it is sufficient to note that there is nothing within the functionalist perspective alone that can adjudicate on the extent of the realizer of a mental state. As Shapiro argues, functionalism is the wrong perspective from which to judge the merits of the EM theorists' arguments.

Even if Clark could secure the extension of the realizer of Otto's belief in manner that he requires it is not clear that even this would be sufficient to secure the extension of Otto's belief into the world. This is because functionalism, quite apart from being ill-equipped to settle the boundary problem with regard the realization of a mental state, is inadequate insofar as a complete characterisation of the nature of the genuinely mental is concerned.

Although functionalism in many ways improves on behaviourism it does not seem clear that functionalism is necessarily better placed to make sense of what are commonly taken to be the special and essential properties of the mental. As we saw in earlier sections a pure commonsense functionalist account of mental states specifies that what it is to be a certain mental state is nothing more than to fulfil a particular causal role. But specifying the functional role alone is not considered sufficient to explain the paradigmatically mental properties, viz. qualitative experience, and intentionality.

For example, if qualia are wholly definitive of consciousness and wholly non-functional then there will be no extension of consciousness as based on the extension of whatever realizes functional role. Consider the possible functional definition of pain that we considered briefly earlier in the chapter. If what it is to be in pain is something like to believe that as a result of injury, there is something wrong with the body, to desire to no longer be in that state, to move away from the cause of the injury so as to minimize damage, and to moan and possibly cry, then on a purely functionalist account of mental states it is possible to be in the functional state of having pain without the usual attendant qualitative content—pain without the sensation of pain. It seems, then, that functionalism on its own is not sufficient to account for the qualitative content of mental states, which raises the question of in what sense functionalism is correctly characterising mental states at all (cf. Block 1980).

> It does not…seem entirely unreasonable to suggest that nothing would be a token of the type "pain state" unless it felt like a pain, and that this would be true even if it were connected to all the other psychological states of the organism in whatever ways pains are. (Block & Fodor 1972: 172)

But the arguments of the EM theorists that we have been looking at do not argue for the extension of qualitative experiences into the world. And, indeed, in Clark and Chalmers' original paper it is explicitly stated that this variety of mental state may not be amenable to EMT in the same way as propositional attitudes. Furthermore, it is plausible to argue that there is no qualitative content for at least some propositional attitudes, particularly beliefs: "beliefs are not *inner feelings* whose causal links are available to introspection" (Jackson & Pettit 1988: 387, emphasis mine).

Thus, the problem of accounting for qualitative content may not be one that touches the propositional attitudes that the EM theorist seeks to extend. Be that as it may, qualitative content is just one of the properties that are paradigmatic of mental states. The other is intentional content, and this is also argued to be problematic for the purely functionalist account of mental states.

Anyone who rejects the possibility of a purely syntactic theory of beliefs will reject the idea that there is nothing more to believing that *the Museum of Modern Art is on 53rd Street* other than fulfilling the requisite causal role. If content is wholly definitive of propositional attitudes and will not functionalise then there will be no extension of content as based on the extension of whatever realizes functional role. And if propositional attitudes can be partly defined in terms of functional role and if part of the functional aspects extend then we still don't have extended mental states since what is definitive of mental states, what distinguishes them from the rest of the world would still be internal.

Clark is happy to stick with wholly functional characterisations of mental states when presenting his EMT but when challenged he argues that he can allow that non-derived content is the mark of the mental and that it is internal and still have his extension of mental states into the world. This is, *prima facie*, a strange position to adopt given his claims active externalism should be viewed as: "more like an environmentally extended case of narrow content than a case of broad content" (Clark, forthcoming). It is also strange because at no stage in his argument does Clark present a case for extended narrow content. The only way to reconcile this statement with the arguments Clark presents would be to say that he thinks a purely functional account of narrow content is possible. But he gives us no reason to think that this should be the case. And when challenged on the issue of content by Adams and Aizawa he

retreats to the claim that although non-derived content may be internal the functional aspects of mental states extend.

The point is not that functionalism is a dead end with regard to the possibility of giving an account of the nature of mental states, rather, the point is that functionalism *on its own* cannot do justice to the paradigmatically mental properties; content and qualia. As Adams and Aizawa point out, there is more to mentality than passing the Turing test. Thus, as Block (1980) argues, specifying or seeking to account for mental states in terms of causal role alone, as Clark and Chalmers do, leads to liberalism with the attribution of mentality. If fulfilling a certain causal role is all that there is to being in a particular mental state then mentality will be attributed to things/processes/states that are undeserving of it. This is why functionalism is thought to require supplementation with a representational theory of mind.

Thus, even if the EM theorist could secure the extension of the realizer into the world such that the sentence in Otto's notebook was not just an enabler of the realizer of his mental state but a constituent part of it, it is not clear that the extension of the realizer would secure an extension of the mental in any interesting sense unless the extension of the realizer entailed an extension of content. But if non-derived content is the mark of the mental and if formal, physical properties alone and the relations between them are not sufficient so far as a complete characterisation of mental states is concerned, since they cannot secure the contentful properties of mental states, then it is not clear how the extension of these formal, physical properties could be sufficient to secure an extension of the mental in any interesting sense. The EM theorist owes us a theory of content then, and not only that, he owes us an explanation of how this (narrow) content can be said to extend into the world. Without this, arguments for the extension of the realizer of Otto's belief into the world

are arguments for the extension of "mere processing" and by Clark and Chalmers' own account, not sufficient to secure an extension of the mind into the world.

In the previous section we noted Rupert's characterisation of the EM theorist's argument as based on functionalism.

> P1: A mental state of kind F is realized by whatever physical state plays the functional role that is characteristic (or metaphysically individuative) of F.
> P2: Some realizations of functional mental state kinds have physical components external to the organism.
> P3: A mental state extends to or includes all components of its realization.
> C: Therefore, some mental states extend beyond the boundaries of the organism. (2004: 34)

Rupert argues that neither the commonsense characterisation of mental states nor a version of psychofunctionalism can be made to accord with premise two of this argument. He may well be correct in this but, nonetheless, my argument is that the EM theorist faces a real problem with regard to premise 3. EM theorists focus exclusively on extending the formal, physical properties of the realizer of mental states but this will not be enough to secure an extension of the mental in any interesting sense.

Without the addition of a representational theory of mind to Clark's extended functionalism we have no reason to regard the specified extended realization as capturing what is essential to the mental state—its content—and with the representational theory of mind on board, we have no reason to think that the realization extends to include the notebook. At

best, then, the sentence in Otto's notebook has a special enabling role to play in the realization of Otto's occurrent belief. Without the sentence in the notebook there will be no (53rd Street directed) action, but this is the same thing as saying that there would be no blood pumped throughout the body without the contribution of the blood vessels, this does not entail that the blood vessels are constitutive parts of the realizer of a heart. The functionalist characterisation of a heart does not provide the means by which to settle the extent of the heart's realizer. Likewise, a pure functionalist characterisation of a belief *that the Museum of Modern Art is on 53rd Street* does not provide the means by which to settle the extent of the state's realizer. Here, other factors, such as the requirement to provide an account of the non-derived content that individuates the mental state, would seem to settle the matter in the internalists favour. Thus, there is a dilemma for the EM theorist; either his claim is that all there is to being a mental state is that it fulfils a certain functional role, in which case he is endorsing a form of functionalism that is overly liberal in awarding mentality to states that may not possess non-derived content; or he is endorsing non-derived content as the Mark of the Mental, in which case he has provided no reason to think that mental states extend.

EMT and Cognitive Processes

The EM theorist could retreat at this point from claims made regarding the extension of mental states into the world to the ones concerning the extension of cognitive processes only. But this strategy will only work provided that Adams and Aizawa are wrong in arguing that non-derived content is also the mark of the cognitive, in which case we will have no reason to regard it as moving beyond the extension of 'mere processing' to the extension of the genuinely mental.

What Clark and Chalmers may well have demonstrated with their Otto example is a case of cognition extending, which means that it may well be correct to say that when Otto is accessing his notebook (automatically and transparently we can suppose) that he is recalling the location of the Museum of Modern Art. And, indeed, the notebook may have a special role to play in driving Otto's action in this case. But this is not the same thing as arguing that Otto's belief extends into the world. If we can distinguish between cognitive processing and mental states, as Clark and Chalmers claim, then a demonstration that Otto's cognitive processing extends does not equate to a demonstration that his mental state extends. Rowlands seems to appreciate this since his primary focus (1999, 2003, 2009) is on the extension of cognitive processing into the world, admitting that these arguments work only so long as cognitive processes are specified non-intentionally.

Conclusion

Clark seeks to extend mental states into the world and takes his paradigm mental states to be propositional attitudes. But since non-derived content is the definitive feature of propositional attitudes and Clark concedes that non-derived content is internal, his argument fails to conclusively prove that mental states extend. At best, given the strategy that he adopts, Clark can argue for the extension of the 'special enablers' of mental states into the world. But this falls short of extending mental states into the world in any interesting sense and can be described, instead, in terms of EMMT. Within the EMMT framework it may be legitimate to argue for the extension of cognitive processes into the world provided that those cognitive processes are specified non-intentionally. But as we have seen, and as Clark admits, the extension of cognition is not the same thing as the extension of mental states, which is necessary for the extension of mind.

If Clark was interested in arguing for a genuinely interesting and challenging version of EMT then instead of accepting non-derived intentionality as the mark of the mental he could have challenged this view and attempted to provide a new mark of the mental. By not presenting such a challenge Clark trivialises his argument. In this regard Clark goes too far in arguing for the extension of mental states into the world because he does not go far enough in his rejection of the typical/traditional view of the nature of mental states that he seeks to challenge. A genuine challenge to this view could take the form of providing a new mark of the mental by rejecting the representational theory of mind. This is what I explore in the next chapter.

Chapter 6

Introduction

As currently formulated, EMT fails to decisively show that mental states extend into the world. Relying on functionalist principles means Clark provides arguments to conclusively demonstrate only that the computational processes supporting cognition and mental state formation extend. The only other possible move for defenders of EMT to make (apart from extending representations with non-derived content) would be to argue that although not all aspects of mind extend it is at least possible that the functional aspects of non-biologically basic mental states sometimes extend, in certain special cases, when they are appropriately related to internal representations with non-derived content.

In this chapter I give a brief diagnosis of what I take to be a strategic flaw in Clark's approach. Because of his willingness to accept the standard representational view of intentional states he has missed an opportunity to defend EMT on stronger grounds. As shown by the preceding chapters it is because Clark allows that internalists may be right about the defining properties of intentional states that his arguments for EMT fail to convince. This suggests that the strongest move in defending EMT would be to question standard thinking about what is definitive of intentional states. With this in mind I examine the anti-representationalist stance of the enactivist approach to cognition and mentality, demonstrating that such an approach may hold promise with regard to formulating a new mark of the mental and, consequently, of providing a more secure theoretical basis for a genuine EMT.

In chapter four we saw that the constitution versus mere enabling issue is crucial for the EM theorist. One cannot adjudicate on this issue—and therefore on whether EMT is really only EMMT—without appeal to an independently motivated mark of the mental. One cannot specify the location of a mental state without prior specification of what a mental state is. As Di Paolo puts it: "Before asking where it is we *must* first say what it is" (2009: 10). Despite this Clark makes no mention of a mark of the mental in presenting his positive thesis.

In comparison, Rowlands, arguing for the extension of cognitive processes into the world, offers the following mark of the cognitive:

> A process P is a cognitive process if and only if (i) P is essential to the accomplishing of a cognitive task T, and (ii) P involves operations on information bearing structures, where information carried by such structures is relevant to task T. (1999: 103)

We can debate whether or not this mark of the cognitive is up to the job but at least it is recognised that specification of what cognitive processes are is required for specification of their location. The issue is recognised as being sufficiently important for Rowlands (2009a) to update his mark of the cognitive, attempting a more fine-grained specification of cognitive processes as follows:

> A process P is a cognitive process if and only if:
>
> 1) P involves information processing—the manipulation and transformation of information-bearing structures.
> 2) This information processing has the proper function of making available either to the subject or to subsequent processing

operations information that was (or would have been) prior to (or without) this processing, unavailable.

3) This information is made available by way of the production, in the subject of P, of a representational state.
4) P is a process that belongs to the subject of that representational state.

Similarly, Hurley (1998b), in presenting her case for the extension of the vehicles of the unity of consciousness into the world, recognises that she requires a specification of vehicles that makes a discriminating appeal to causal spread. Thus, she offers a mark of the vehicle (as we saw in chapter three) that appeals to its duplicability in counterfactually altered environments.

But the extension of vehicles does not necessarily entail the extension of mental states. Likewise, even if Rowlands' mark of the cognitive allows the extension of cognitive processes, this does not mean that it allows the extension of mental states. Although Rowland's updated mark of the cognitive might allow the extension of cognitive processes that might enable mental states, what is definitive of the mental is taken to be internal: "cognitive processes always contain a non-eliminable internal element. It is here that we find representational states that possess non-derived content" (Rowlands 2009a: 13).

Clark does not offer a mark of the mental for the intentional states that he wishes to extend. By relying on the parity principle and appeals to intuition about what should count as mental in order to make his case Clark seems quite happy to let prevailing views dictate what the nature of a mental state is, seeking only to challenge the boundaries that are prescribed. But, as we will see, this is problematic for his stated aims.

The issue of the mark of the mental first raises its head in the EMT debate when (as per chapter four) Adams and Aizawa (2001) propose that non-derived content is the mark of the *cognitive*. But since we can distinguish a mental state from a cognitive process a mark of the cognitive is not the same thing as a mark of the mental. Adams and Aizawa recognise this:

> We assume without argument that the mental is not the same as the cognitive, hence that the mark of the mental is not the same thing as the mark of the cognitive. (2001: 48)

Nonetheless, Adams and Aizawa take their argument to show that the sentences in Otto's notebook do not partially constitute Otto's *belief*, which is a mental state. It is more commonly taken, in line with Brentano's thesis, that intentionality, or (as it is typically taken) non-derived content, is the mark of the *mental*. And given that our target here is the extension of mental states, I propose to take non-derived content as the mark of the mental rather than of the cognitive. This allows that cognitive processes might extend while mental states remain internal. But Clark thinks that he can allow non-derived content as the mark of the mental and still argue for the extension of mental states.

Clark offers an initial challenge to the idea of non-derived content as the mark of the mental, arguing, in line with the parity principle, that we would have no problem allowing the derived content involved in bitmapped storage as constitutive of a mental state if it occurred in the head:

> Surely, I argued, we would have no hesitation in embracing that kind of bitmapped storage, even prior to an act of retrieval, as part and parcel of the Martian cognitive equipment … If, courtesy of our common-sense

> psychological intuitions, we accept this aspect of Martian memory into the cognitive fold, surely only skin-and-skull-based prejudice stops us from extending the same courtesy to Otto. (Clark 2008b: 91)

But if non-derived content is indeed the mark of the mental then it is not our common-sense intuitions that (should) award mental status. If non-derived content is the mark of the mental and if this rules out derived content from forming a constitutive part of a mental state then it doesn't matter where the derived content is located, it cannot count as a constitutive part of a mental state. Derived content, regardless of its location, requires non-derived content; this is where it inherits its contentful properties from. And as the EM theorists argue, just because a process or state is internal doesn't accord it mental status. Parity (or 'fair treatment' as Sprevak would put it) cuts both ways: if a process or state occurs within the head to which we would deny mental status had that process or state been in the world, then that process or state is not a constitutive part of a mental state (Coleman, forthcoming). If non-derived content is the mark of the mental then the bit-mapped aspect of Martian memory is not accepted as constitutive of his beliefs and we do not award mental status to Otto's notebook on precisely these same grounds.

Of course the question arises here as to what we require from a mark of the mental. Adams and Aizawa (2001), in specifying non-derived content as the mark of the cognitive, are offering a *necessary* condition on a state or process being cognitive and admit that not every part of every cognitive process need possess non-derived content. A mark of the mental that specifies non-derived content only as a necessary condition for awarding mental status would (as we saw in chapter four) certainly *seem* to open the door for Clark with regard to the extension of mental states into the world. If insisting that non-derived content is the mark of the mental merely provides a necessary condition for a state or process as

a whole to count as mental then Clark can argue that, for example, the functional aspects of states that deal with only derived content might be partially constitutive of mental states after all. This would require that those aspects dealing only with derived content are related in the right way to those with non-derived representational content:

> One way to understand this proposal is as insisting that all that matters, for some conventional encoding to count as the vehicle of a dispositional belief, is that it be appropriately linked, at run-time, to representations whose content is (as Adams and Aizawa insist) intrinsic. Such linking can be achieved for conventionally formatted representations both inside and outside the head. (Clark 2005: 5)

In this way Clark thinks that he can allow that non-derived content is the mark of the mental and still have his extension of dispositional beliefs into the world. This requires only that the processing dealing with derived content is specially related in the right way to internal representational states possessing non-derived content. Thus, Clark argues, if we accept that non-derived content offers only a necessary condition on a state being mental then we have no good reason to reject part of a state—that does not itself display this mark of the mental—as a constitutive part of a mental state, regardless of its location, provided that it is coupled in the right way to a internal representation with non-derived content. But an internalist is likely to respond that if non-derived content is the mark of the mental then Clark has given no good reason to think that the mental state itself has extended rather than the mere enabler of that mental state.

What Clark's argument trades on is the idea that there may be elements that are characteristic of a mental state other than its non-derived content. He claims that accepting non-derived content as the mark of the mental (or at least not explicitly rejecting it) allows for the possibility of the

extension of these other elements even if the representational aspect, the non-derived content, remains internal. But, as we saw in the previous chapter this is to argue only for the extension of the causal/ computational/syntactical aspects and not the representational aspects, not the non-derived content that is taken to demarcate the mental from the non-mental.

Furthermore, if these causal/computational/syntactical elements are to count as partially constitutive of mentality then it is difficult to see how we can draw a clear and non-arbitrary distinction between those that are internal and those that are external. It seems there is no principled way to settle the boundary dispute since there is nothing else to appeal to. Having agreed on a mark of the mental there seems nothing else to call on that would allow us to say which functional states are genuinely constitutive of mentality and which are merely causal or enabling of mentality.

Clark does offer some extra conditions on an external resource forming part of a mental state or cognitive process. These are introduced in order to make a discriminating appeal to causal spread and avoid the problem of bloat. But these extra conditions apply *only to external* resources and are there to ensure that not just anything that is causally linked to a representational state with non-derived content counts as mental.[35] But the internalist will object that causation is not the same thing as constitution and that a mark of the mental should apply to all putative mental states equally, be they internal or external. Thus, genuine mental states shouldn't need to fulfil the specified criteria in addition to the mark of the mental just to avoid the problem of bloat. Better to abide by the

[35] Aizawa, ms, argues that these conditions are so restrictive that if they were applied to internal states then they would rule clear cases of cognitive processing as non-cognitive, and similarly rule out clear cases of belief.

mark of the mental and restrict mental states to internal states they will argue. Although Clark's extra conditions might avoid the problem of bloat for what can count as an enabler of a mental state it does not settle the constitution versus enabling issue.

If non-derived content is *the* wholly definitive mark of the mental—such that it alone distinguishes the mental from the non-mental—then the extension of states and processes that do not display this mark (regardless of any extra conditions placed on them) will fail to secure the extension of the mental in any interesting sense. For those who accept the above, they will argue that at best, Clark has an argument for the enablers or facilitators of mental states—he has an argument for EMMT as opposed to EMT.

Clark's strategy provides no decisive way of adjudicating on the EMT debate, because it will not settle the issue of what is constitutive of a mental state as distinct from what merely enables a mental state. Agreeing that non-derived content is only a necessary condition on a state's being mental and not definitive of mentality is not enough to settle the important issues regard the metaphysical extent of mind. The root problem is that there is nothing left to appeal to that would decisively enable us to choose between accepting an embedded mind hypothesis, compatible with a non-extended mind, or EMT. To break the deadlock a mark of the mental that offers more than a necessary condition on a state's being mental is needed. We require of a mark of the mental that could definitively settle the constitution versus enabling issue.

Extended Representations

It is instructive to consider Clark's apparent unwillingness to tackle the issue of the mark of the mental head-on. This may be representative of

what Di Paolo (2009) flags as a tendency by EM theorists to simply accept the orthodoxy on what cognition and mentality are and seek to extend cognition and mental states according to this definition. And the orthodoxy is representationalism:

> While there is no agreement about the best version of the representational theory of mind and its vehicles, it is widely assumed that some version or other is non-negotiable. In this respect, an unquestioned commitment to representationalism frames contemporary debates about cognition in precisely the way that a similar commitment to some 'theory of ideas' or other framed the debates that raged between empiricists and rationalists centuries ago. (Hutto 2008: 420)

Clark seems willing (following the initial resistance outlined above) to accept non-derived content as a necessary condition on a state's being mental because he thinks he can preserve the extension of mental states into the world even while making this allowance. On this view derived content, like that found in a sentence in Otto's notebook, can form a constitutive part of a mental state provided that, to use Clark's words, is 'appropriately linked' or 'specially related' to a representation with non-derived content.

But this invites the response that what is truly definitive of the mental remains internal, particularly if we require our mark of the mental to specify more than a mere necessary condition on a state's being mental such that it can settle the constitution versus enabling issue. Thus, Clark's unwillingness to challenge the representational view of mind and in particular non-derived content as the mark of the mental prevents him from developing a logically compelling argument concerning the metaphysical extent of minds. A possible move towards the development

of such an argument might be for the EM theorist to argue for the extension of non-derived representations into the world.

Rowlands (2006) attempts to extend representations into the world in the form of actions. If such an endeavour can succeed then perhaps a stronger representationalist version of an argument for EMT might be fashioned. Rowlands holds that the concept of representation has suffered as a result of being assimilated to the category of the word and he suggests a re-evaluation of representation; arguing that pre-intentional actions, or deeds, can be seen as representational. Deeds include positioning your fingers in order to catch a ball coming towards you at speed and the movement of your fingers when playing the piano, for example (Ibid.). Deeds are seen as being pre-intentional because they involve "an array of on-line, feedback modulated adjustments that take place *below the level of the intention*, but collectively promote the satisfaction of the antecedent intention" (Ibid: 103) and also, crucially, because "the direct antecedents of these deeds are not themselves representational or intentional states" (Ibid: 104).

It is beyond the scope of this work to go into a detailed examination of Rowland's arguments here but it suffices for our purposes to list what Rowlands thinks are the characteristics of these pre-intentional deeds such that they are representational, they:

1. Carry information about x (e.g. the trajectory of the ball).
2. Track x or function in a way that allows the subject to accomplish something in virtue of tracking x.
3. Can misrepresent.
4. Are decoupleable from x (e.g. I can 'practice' catching the ball or rehearse a catch in the absence of any ball) (Co-opted from Gallagher 2008).

By way of contrast, the classic concept of a representation has it that:

1. Representation is internal (image, symbol, neural configuration).
2. Representation has duration (it's a discrete identifiable thing).
3. Representation bears content that is external to itself (it refers to or is about something other than itself).
4. Representation requires interpretation – its meaning gets fixed in context.
5. Representation is passive (it is produced enacted, called forth by some particular situation; or we do something with it).
6. Representation is decoupleable from its current context. (Gallagher 2008: 351-352).

Rowlands' account of pre-intentional deeds as representations seeks to challenge conditions one to five of this classic concept of representations, hence Rowlands allows that not all of the most basic kind of representations need be internal. Given that Rowlands' account of representations as deeds rejects so much of what is commonly taken to be definitive of representations we can legitimately question to what extent Rowlands' deeds are genuine representations at all, although we might allow the weaker claim that they are *representational* in some sense.

Gallagher (Ibid.) makes exactly this challenge, arguing that this minimalist conception of representations no longer conforms to the criteria that would make it a representation. If so, to think of deeds as representations requires surrendering many of the core criteria that must normally be satisfied for something to count as being a representation at all. Hence it is most likely that deeds will not be counted among the ranks of genuine representations. And if possession of genuine representational content is taken to be the mark of the mental then it doesn't look like Clark's version of EMT would be aided by appeal to Rowland's proposal.

Furthermore, it is difficult to see how pre-intentional deeds might stand in relation to the kinds of propositional attitudes that Clark makes his focus. Pre-intentional acts require something like the subpersonalisation of action that we mentioned in chapters one and three. But it is difficult to see how Otto's action of accessing his notebook could be subpersonalised in the same way as the access movements of a commissurotomy patient might. Even so, if Otto's accessing of his notebook only required an antecedent subintentional state like a deed to initiate that action, we would still be owed an account of how Otto's action could be representational. And even then it would still seem that explanation of Otto's subsequent action as based on the information in the notebook would require a subsequent believed internal representation of the sentence in the notebook. Attempting to salvage EMT by appeal to extended representations in the form of deeds is going to be problematic. This suggests that the cleanest move in securing a genuine EMT would be a more radical and wholesale rejection of the representational view of mind.

Enactivism

Given the difficulties outlined above, a more promising move for developing an interesting version of EMT that secures the idea that mental states genuinely range into the world is to reject the representational view of mind. Enactivist approaches to mind and cognition do exactly this and here I explore the possibility of enactivism providing a different theoretical basis from which to argue for a version of EMT by providing an alternative mark of the mental.

Much (though certainly not all) of the motivation for the enactivist view of mind can be said to come from the failure of classical AI to adequately model human intelligence. Two principle features of classical AI were:

1. Adherence to the rules and representation model of mind.
2. A top-down approach to defining cognitive processes and mental states.

The rules and representations model suggests "that it is possible to simulate intelligence in a particular 'world' by manipulating a set of symbols that represent that 'world'" (Lenat & Guha 1990: 147). Designing an artificially intelligent system on this view requires representing various aspects of the world and articulating appropriate rules to manipulate these representations such that the appropriate outputs will be produced. Also, Classical AI takes a very particular stance on the kind of processes that are definitive of intelligence. Cognition is seen as a high-level phenomenon that is highly abstract and disembodied. Implementing chess playing programs and medical diagnosis programs are seen as examples of the proper targets for AI on this approach.

This led to the creation of expert systems that equalled and sometimes surpassed humans at the specific tasks they were designed to perform. But the problem with these expert systems is that they are too specialised; once removed from their own fields of expertise they are woefully inept. Expert systems are extremely susceptible to failure when confronted with novel situations outside of their area of expertise and are, consequently, said to be brittle. Classical AI argues that the way to deal with new situations is "by finding some related case and propagating the differences to this new one" (Ibid: 3). And the way to do this is to ensure that there is large base of knowledge to consult. Thus: "the mattress in the road to AI is lack of knowledge, and the anti-mattress is knowledge" (Ibid: 4).

So the reason that the expert systems developed by the Classical AI approach are brittle is not, on this view, because the rules and representation model of intelligence is wrong but rather because the

expert systems are lacking in knowledge. Expert systems can achieve their specific tasks without 'knowing' anything about the area outside their task, and without 'knowing' very much about the task itself. To overcome the brittleness of the expert system Classical AI postulates that we need to design a system that knows lots of things about the world: to broaden their competence and ability to deal with novel situations expert systems need more propositional knowledge, more internal representations.

It is along this line that the CYC project was undertaken by Douglas Lenat. The aim is to encode in the system's knowledge-base as much as possible of our consensus reality: "the millions of things that we all know and that we assume everyone else knows" (Ibid: 4). The hope is that there is a threshold that this system will cross at which point it will be able to assist with its own programming and will, ultimately, become capable of learning for itself. Despite some commercial success, in the 26 years since its inception, the CYC project has not achieved this. Pronouncements that CYC has succeeded in modelling human intelligence have not been forthcoming.

The key to success in Classical AI is seen as representing the world completely and explicitly and this, suggests Brooks (1991) is precisely the issue on which AI has foundered. He recommends an alternative approach to artificial intelligence whereby instead of trying to explicitly represent as much of the world as possible:

- We must incrementally build up the capabilities of intelligent systems, having complete systems at each step of the way and thus automatically ensure that the pieces and their interfaces are valid.

- At each step we should build complete intelligent systems that we let loose in the real world with real sensing and real action. Anything less provides a candidate with which we can delude ourselves. (Ibid: 139)

Thus, Brooks rejects the top-down approach to cognition advocated by Classical AI and advocates, instead, a bottom up approach whereby: "mobility, acute vision and the ability to carry out survival-related tasks in a dynamic environment provide a necessary basis for the development of true intelligence" (Ibid: 140). In adopting this alternative approach Brooks created a series of autonomous mobile robots and in the process reached what he calls an unexpected conclusion (C) and a radical hypothesis (H):

(C) When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model.
(H) Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems. (Ibid: 139)

The failure of Classical AI to effectively model human level intelligence, coupled with his own success in designing mobile, autonomous robots also leads Brooks to shun the rules and representations model of mind. In this way Brooks rejects the view that the sorts of states and processes that are definitive of human mentality are the high level, abstract, representational states and processes that are the focus of classical AI. The true character of intelligence is to be found he suggests in explorations of how "whole, physically embodied agents, including nonhuman animals, achieve real-time sensorimotor control in dynamic, sometimes unforgiving environments" (Wheeler 1995: 1). In rejecting the rules and representations model of Classical AI and advocating a bottom-up approach to explaining cognition Brooks can be seen as an important forerunner to the enactivist approach.

Enactivist approaches to mind similarly reject the representationalist view of mind. Representationalism begins from the positing of a fundamental gap between mind and world that requires the postulation of mental representations to explain possibility of engagement with the world: "There is a gap between the mind and the world, and (as far as anybody knows) you need to posit internal representations if you are to have a hope of getting across it" (Fodor 2009). The enactivist approach rejects the positing of this gap and instead takes as its starting point the organism *in the world* rather than in isolation from it. So, in rejecting the representationalist approach enactivism also endorses a bottom-up approach to cognition and mentality whereby "Abilities are prior to theories ... Competence is prior to content … [and] knowing how is the paradigm cognitive state and it is prior to knowing that" (Fodor 2008: 10). This gives us a clear picture of the explanatory programme of the enactivist perspective: mentality emerges from the self-organizing activity of the organism through its extended interactions with the world.

Enactive Perception

To make the distinction between the enactivist and representational approaches more concrete, I will very briefly contrast an enactivist account of visual perception with the standard representationalist account. Traditional accounts of perception attempt to explain our visual experience of the world in terms of the creation or activation of internal representations of that world. So, your experience of a shiny, green apple is to be explained, on this view, by the creation or activation of an internal representation, the components of which represent each of these features of your experience, i.e. the shininess of the apple, its greenness, and its roundness.[36] And, given that our visual experience of the world is

[36] This probably goes without saying but to be clear; requiring that a representation has features that correspond to the different features of your experience is not to say that the representation must be similar in any way to the experience (although it might be).

rich, detailed and vivid this is to be explained in terms of a rich, detailed and vivid internal visual representation.

But experiments carried out by Kevin O'Regan in particular (e.g. O'Regan, Rensink & Clark 1999; O'Regan & Noë 2000, 2001) seem to call into doubt this rich and detailed character of our visual experience. In fact, O'Regan's experimental work seems to reveal a certain lack of detail in the human visual experience that can easily become manifest under experimental conditions. We are, it seems, subject to a phenomenon called change blindness:

> When a few small, high contrast shapes are briefly spattered over a picture, like mudsplashes on a car windshield, very large changes can simultaneously be made in the scene without these being noticed. This occurs even when the mudsplashes do not in any way cover or obscure the changes. (O'Regan, Rensink & Clark 1999)

So, in these experiments subjects are shown a visual scene in which large scale changes are made repeatedly on a cyclical basis. These are changes that would easily be noticed by the subject under normal conditions but when a distracting stimulus is presented (such as mud splashes) at the same time as the change then subjects have great difficulty in seeing the change being made. This will be the case even though the changes might occur in full view and are not obscured by the distracting stimulus. Some have argued that these results cast doubt on the representational account of visual perception, since on this account:

> [A]ll that would be required to notice a change in such a scene would be to compare one's current visual impressions with the activated representation; when and how the discrepancies between the former and

---

There is no requirement for a representation of something green and shiny to itself be green and shiny, only that the representation must represent greeness and shininess.

> the latter arose would be irrelevant. Thus, it is argued, the change blindness results support the claim that there is at least no complex and detailed internal representation. We do not notice even significant changes in a scene because we have no internal template against which to measure or compare them.[37] (Rowlands 2006: 69-70)

The results of the change blindness experiments, among other things, have led some to label our visual experience of the world as a 'grand illusion' (e.g. Dennett, 1991, 1992, 1998; O'Regan, 1992; Rensink, O'Regan & Clark, 1997).

> We believe that we see a complete, dynamic picture of a stable, uniformly detailed and colourful world, but [o]ur stable visual world may be constructed out of a brief retinal image and a very sketchy, higher-level representation along with a pop out mechanism to redirect attention. The richness of our visual world is, to this extent, an illusion. (Blackmore et al. 1995: 1075)

On this view we are mistaken about the nature of our visual experience. We are mistaken in thinking that our experience of the world is rich and detailed, presumably because it is not supported by a sufficiently rich and detailed internal representation. But Noë argues, contra Blackmore, that: "It just is not the case that we, normal perceivers, believe we see a complete, dynamic picture of a stable, uniformly detailed and colourful world" (2002: 6). Instead, Noë claims that it seems to us as if we have perceptual access to a *world* that is stable, uniformly detailed and colourful. To think otherwise is to hold that perceivers believe what Noë calls the snapshot conception of visual experience.

[37] I do not critique the representational view here. I merely use the change blindness experiments as a means of contrasting the approaches of the representationalist and enactivist positions.

> According to a conception of visual experience that has been widely held by perceptual theorists, you open your eyes and — presto! — you enjoy a richly detailed picture-like experience of the world, one that represents the world in sharp focus, uniform detail and high resolution from the centre out to the periphery. Let us call this the snapshot conception of experience. (Noë 2002: 2)

Noë claims that much empirical research into visual experience takes this snapshot conception as its starting point. When you combine this conception of visual experience with the reality of the limitations of our perceptual apparatus, i.e.:

> [T]here are two retinal images, not one, and they are distorted, tiny and upside-down … In addition, the resolving power of the eye is limited and nonuniform; outside the high-resolution foveal region, the retina is nearly colour-blind and its powers of discrimination are severely limited. On top of this, the eye is in nearly constant motion, saccading from point to point in the visual field three or four times a second. As a result of saccadic suppression, the data made available to the retina takes the form of a succession of alternating snapshots and grey-outs. (Ibid: 2)

…then it is difficult to see how one could do anything other than posit the existence of rich and detailed internal representations of the world as a means of explaining visual perception. But, crucially, as we saw above, Noë argues that the snapshot conception is mistaken. It is the world that is rich and detailed and we have access to all of this richness and detail but not all at once. We perceive a rich and detailed visual world thanks to our capacity to move our heads and our eyes, to shift our attention, to act in the world.

Thus, the enactivist approach rejects talk of a grand illusion. We experience a rich and detailed visual world not because we form rich and detailed internal representations of the world but because we have an ability to instantly attend to any aspect of the scene we choose and we have implicit knowledge of how that scene will change depending on how we explore it. A tomato that is partially obscured by a pepper shaker is phenomenologically present to the observer, rather than experienced as distinct tomato parts, because:

> Our perceptual sense of the tomato's wholeness—of its volume and backside, and so forth—consists in our implicit understanding (our expectation) that the movements of our body to the left or right, say, will bring further bits of the tomato into view. Our relation to the unseen bits of the tomato is mediated by patterns of sensorimotor contingency. (Noë 2004: 63)

Sensorimotor contingencies, or dependencies, simply refer to how the sensory stimulation changes, or is liable to change, given certain movements or environmental changes. So, for the enactivist, the perceptual experience of a rich and detailed visual world is the result of activity that goes on in the world: "seeing is a skilful activity whereby one explores the world, drawing on one's mastery of the relevant laws of sensorimotor contingency" (O'Regan & Noë 2001: 966).

The diachronic engagement of the organism with the world is, for the enactivist, central to an explanation of the-what-it-is-like of token visual experiences. This might be compatible with (though would not presuppose) a representational account of vision, so long as the bodily movement and patterns of sensorimotor contingencies had only an instrumental role to play in perception. This has echoes of Kirsch and Maglio's claims regarding epistemic actions that we examined in chapter

one. They recognise the important causal role that epistemic actions have to play in the transformation of the cognitive task facing a subject, but do not make the stronger claim that epistemic actions can be constitutive of cognitive process and mental states. Similarly, a proponent of the representationalist account of visual perception might allow that patterns of sensorimotor contingencies have a causal role to play in perception by productively reconfiguring the tasks performed in the brain, though they are not constitutive of the perceptual experience itself.

Some theorists move for a stronger claim than mere instrumental dependence of perception on action and argue that the temporally extended use of bodily movement and knowledge of various sensorimotor contingencies have more than a mere causal or explanatory, role to play in perception. On this view the what-it-is-like of perceptual experience can depend non-instrumentally, or constitutively, on bodily movement and implicit knowledge of sensorimotor contingencies (Noë 2004, Rowlands 2003).

But even this stronger claim is compatible with endorsement of the representational view of mind. And, indeed, although the principal SMC (sensorimotor contingency) advocates, like Noë and O'Regan, reject the idea that we need 'rich and detailed' representations to allow us to form a picture or snapshot of the world in perception, they do not surrender the idea that representations are required for perception or cognition more generally. Indeed, sometimes they are explicit about the requirement for representations:

> [F]or perceptual sensation to constitute experience – that is, for it to have *genuine representational content* – the perceiver must possess and make use of sensorimotor knowledge. (Noë 2004: 17, emphasis mine)

In addition to concerns about requirements for 'genuine representational content' there is a worry raised by Hutto (2005) about the mediating knowledge appealed to in an SMC account of perception:

> The central idea of our new approach is that vision is a mode of exploration of the world that is *mediated by knowledge* of what we call sensorimotor contingencies. (O'Regan & Noë 2001: 940, emphasis mine)

> [W]e sought to explain phenomenal consciousness in terms of forms of cognition—for instance, knowledge of patterns of sensorimotor dependence. (Noë 2004: 228)

If these appeals to knowledge are understood in rules and representations terms then this commits SMC theorists to a conservative version of enactivism at best. In contrast, radical enactivists reject the cognitivist assumptions of representationalists and conservative enactivists. They deny:

1) That the character of experience involves or is exhausted by intentional content (narrow or wide);
2) That the relevant forms of embodied activity for necessary for perception are based on knowledge operative at either the personal or subpersonal level.

Radical enactivists argue that rejecting the requirement for a static inner representation and thinking of perception as a kind of exploratory activity that is extended over time means that:

> [T]here is no need to introduce 'knowledge' as a kind of bond that holds together various percepts in order to explain phenomena such as perceptual presence. (Hutto 2005)

Retaining the requirements for representations of some sort and/or appealing to mediating knowledge in an enactivist account of perception or cognition is to risk run the risk of falling into conservative thinking. It is not clear that such an account would be properly enactivist in line with the enactivism first proposed by Varela, Thompson and Rosch (1991). On the other hand, given its firm anti-representationalist stance, radical enactivism looks promising with regard to specifying an alternative mark of the mental that situates the metaphysical extent of mind firmly in the world—if it can be made to work.

A new Mark of the Mental?

Mind and cognition on the radical enactivist view can be seen as essentially and constitutively body and world-involving (e.g. Thompson 2007). Thus, it is possible to view mentality as necessarily supervening on body and world as well as on the brain, making radical enactivism particularly amenable to the possibility of formulating an alternative and more interesting version of EMT. As Thompson puts it: "the human mind is embodied in our entire organism and embedded in the world, and hence is not reducible to structures inside the head" (2005: 408). But success, in this regard will, of course, be dependent on the capacity of the radical enactivist approach to specify a mark of the mental that locates mind in the world.

The prospects are promising since radical enactivism neutralises many of the objections that face current versions of EMT. Because the starting point for radical enactivism is the organism in the world, mind emerges from the activity of the organism in the world and there is no gap between mind and world that requires the postulation of internal representations. Objections assuming that the mark of the mental must be

unpacked in terms of non-derived representational contents and that, consequently, the genuinely mental must be internal carry no force.

The paradigm mental states cannot be propositional attitudes on this view. The core, low-level, ontologically basic forms of cognitive activity that are required for higher-level cognition are the paradigm of mentality. And given that these core cognitive activities necessarily supervene on brain, body and world, mind is properly seen as *extensive* rather than *extended* (thanks to Dan Hutto for this). Where EMT, as currently formulated, allows only that hybrid cognitive states and processes that involve 'add-ons' like Otto's notebook, might on occasion extend into the world, radical enactivism has it that the mind is essentially world-involving and so does not require extension into the world.

Thus, the distinction between cognitive processes and mental states dissolves, or rather, it doesn't arise. Because the radical enactivist approach does not posit the requirement for internal representations as a starting point and adopts a bottom-up approach to explaining mind, the distinction between mental states and cognitive processes that we saw in previous chapters does not arise. Instead, mentality consists in the active engagement of an organism with the world. Any other higher-level cognitive processes or mental states that we may wish to attribute are based upon and ontologically require these forms of engagement with the world.

If the radical enactivist approach to explaining mentality can be made to work then it seems *prima facie* to offer the possibility of a more thoroughgoing and theoretically well-motivated proposal about the metaphysical extent of minds than is currently on offer. Furthermore it is not susceptible to many of the arguments that dog Clark's position since it does not attempt to take a notion of mind that is primarily internalist

and seek to extend it into the world. But all this is dependent on the radical enactivist approach yielding a mark of the mental. Di Paolo notes that enactivism is working towards a "strong candidate for a widely applicable, non-species-specific, non-bio-chauvinist definition of cognition" (2009: 15) in terms of the normative engagement of an agent with his world. And along these lines he offers a definition of cognition as: "sense-making in the interactive domain" (Ibid.). This requires some fleshing out but this is not the place for that. Instead, I note that we should heed Di Paolo's warning that there is a requirement for further foundational work in enactivism:

> In particular, several essential issues that could serve as a bridge between mind and life (like a proper grounding of teleology and agency) are given scant or null treatment in the primary literature (Ibid: 12).

While yet recognising that if such a definition of cognition is workable then it offers nothing less than an alternative mark of the mental that may form an alternative theoretical basis for a more radical but secure basis for believing that minds have a wide metaphysical extent.

Conclusion

Rick Grush takes issue with what he sees as a growing trend in cognitive science to endorse the view that "the mind is not in the head or that cognition does not require representation, or both" (2003: 53). Clark's EMT seeks to challenge the view that the mind is in the head while remaining wedded to a representationalist orthodoxy. Doing so means that developing decisive arguments for EMT is problematic.

An alternative approach is offered by a radical, non-representationalist version of enactivism. Such versions of enactivism reject the view that cognition requires representation and, by focusing on mind as emerging from the action of agents in the world, challenge the view that the mind is in the head. They adopt a starting point that is much more radical than Clark assumes for EMT. Radical enactivism argues that the mind is not essentially something for representing the outside world. Clark's EMT is much more modest since it:

> [N]eed not deny that the mind is essentially a thinking thing or a representing thing. *It is committed only to the much weaker claim* that some of the thinking, and even the representing, may supervene on activities and encodings that criss-cross brain, body, and world. (Clark 2008b: 149, emphasis mine)

To commit to this weaker claim is to commit to the idea that the mind is fundamentally representational and this leads to difficulties in arguing for anything more than the extension of the mere enablers of mental states. Mind is, on this view, at its core internal and the best the EM theorist can hope for is the *occasional* extension of *some* mental states into the world. Even so, current versions of EMT do not decisively secure the truth of even this claim. Instead, for all that Clark says it remains possible to understand the metaphysics of mind in an internalist way. The cases that Clark cites are compatible with the possibility that minds end at the boundary of the individual even though certain tasks could not be completed without the aid of additional processes going on in the world that *enable* the having of inner mental states of various kinds. This would only show that, sometimes, the machinery of the mind extends as opposed to minds themselves extending. I do not argue here that defenders of EMT have no *possible* way of answering such critics, only that they have no *decisive* way of answering them. My analysis reveals

that acceptance of mental states as essentially representational will provide a serious hindrance to making a convincing and compelling case for the idea the minds are metaphysically extensive.

A better move may be to reject the representational view of mind in line with the radical enactivist approach to mind and cognition. This is approach is still in its infancy but initial progress is promising. In particular, because of its rejection of the representational view and its approach to cognition from the bottom-up perspective radical enactivism can reject propositional attitudes as the paradigmatic mental states and argue that mind emerges from the activity of an agent in the world. In this way enactivism suggests the possibility of forming an alternative mark of the mental that has brain, body and environment as equal partners in constituting the mind.

In the concluding chapter I offer a brief summing up of the principle argument of this work and suggest some possible questions for further study.

Conclusion

The much-needed, original analysis undertaken here highlights numerous problems for the Extended Mind thesis as presented by Clark and Chalmers and subsequently defended by Clark. Their claims for the literal extension of the mind into the world rest on arguments for the extension of mental states into the world. But these arguments must themselves rest upon the provision of a mark of the mental that will settle the issue concerning what constitutes a mental state from what merely enables a mental state.

Clark and Chalmers are not forthcoming with the required mark of the mental, attempting instead to rest the argumentative burden on the much maligned Parity Principle. Consequently, Clark seems happy to accept the standard position; arguing that he can accept non-derived content as the mark of the mental and maintain his argument for the extension of mental states into the world. But it is far from obvious that he can make this position work.

The commonsense functionalist account of mental states that is presupposed by EMT may yield the possibility of extending the syntactic, computational aspects of mental states into the world but if the representational aspects of mental states will not wholly functionalise then the possibility that their truly mental aspects might, even sometimes, extend is questionable. If non-derived content is the defining characteristic of mental states and non-derived content is internal then at most Clark might find the means of defending the claim that certain aspects of mental states might be said to extend on occasion.

We can allow, perhaps, that cognitive processes that do not themselves display the mark of the mental but that support and relate to mentality in

special ways might extend into the world. But if the extension of mental states per se is required for the extension of mind then the extension of cognitive processes alone is not sufficient to secure a genuine EMT. At best, these cognitive processes can be described as operating in relation to internal representations with non-derived content, and making possible the formation of new ones. As such, the extension of cognitive processes secures only the extension of the enablers of mental states (EMMT) and not the constitutive parts of mental states (EMT).

Clark argues that we can count these extended processes as partially constitutive of mental states provided they bear a special relationship to the internal non-derived contents. But it is difficult to see how such a move can be made to work without running into the dangers of bloating or panpsychism. What might do the job in this regard is a supplement to the existing mark of the mental. These additional criteria would need to make a discriminating appeal to causal spread by not accepting just anything that makes a mere causal contribution into the ranks of the mental. But, again, supplementing the mark of the mental in this manner leaves the way open for an internalist to object that since what is definitive of mentality remains internal, what is constitutive of the mental state remains internal.

Alternatively, one could argue that it is possible that representations themselves extend into world. Rowlands (2006) undertakes such an argument but it is problematic for EMT in that a) it is not clear that what he argues for extending constitute genuine representations (as opposed to the weaker claim that they exhibit representationality) and b) whatever content such states might have cannot support the kind of propositional attitudes that Clark takes as paradigmatic of mental states.

In summary, the arguments presented for EMT, as currently formulated, do not definitively secure the extension of mental states into the world and the prospects for securing a genuine EMT by making alterations and additions to the original formulation look deeply problematic and epicylic. A stronger move, then, would be to attempt the formulation of a new mark of the mental.

Brentano proposed intentionality as the mark of the mental and this has overwhelmingly been taken to mean that mental states must bear semantic content. On this view, the paradigm mental states are propositional attitudes—representational states like the belief *that the museum of Modern Art is on 53rd street*. But the enactivist approach to mind rejects the view that mental activity necessarily consists in computational operations performed on internal representations of the world and rejects propositional attitudes as the paradigm mental states.

Instead, enactivism proposes that mind constitutively depends on the activity of the agent in the world. It is through active engagement with environmental features that the mentality of an agent emerges on this view. Such a move may well be compatible with some kind of intentionality being the defining characteristic of mental states but here intentionality is understood in terms of intentional directedness rather than in terms of semantic content. In this way enactivism provides for the *possibility* of an alternative mark of the mental that allows for a genuinely interesting EMT, since the activity that is constitutive of mentality supervenes not only on what goes on in brains, but necessarily on bodily and environmental factors as well. Thus, *if* the enactivist approach can be made to work and *if* it can yield an alternative mark of the mental then it holds the possibility of providing an alternative theoretical basis for a genuinely interesting version of EMT.

The principal conclusion of this work, therefore, is that EMT, as currently formulated, cannot conclusively demonstrate the extension of mental states into the world. Close analysis of the claims of EM theorists and the assumptions on which these claims are based reveals this may be because EMT accepts too much from traditional representationalist views of mind. EMT takes mind as theorised by orthodox representational views and seeks to extend it into world. It's radical only in as much as it seeks to challenge the internalist location claims of traditional views of mind while showing it possible to accept most of their core assumptions. But, as I have argued, current formulations of EMT can unproblematically secure, at best, the extension of cognitive processing into the world and yields only EMMT; the Extended Machinery of Mind Hypothesis. Thus, EMT goes *too far* in arguing for the extension of mental states into the world because *it does not go far enough* in its rejection of the view of mind that it seeks to challenge.

My analysis also paves the way and should provide good grounds and motivation for exploring new and more radical possibilities for a genuine EMT. A more radical EMT, with a new understanding of biologically basic cognition as extensive and not extended would help to deal with Rupert-style concerns that EMT does not provide an interesting basis for re-thinking philosophy of mind (2004). Consider, for example, whether there is any loss of explanatory power should we choose to interpret Otto's case in terms of EMMT rather than EMT. Although current formulations of EMT would, if correct, prompt a re-think of traditional metaphysics about mind, it is not clear that they would prompt a similar re-think in the methodology of cognitive science. Thinking of mental states as embedded rather than extended and of the enablers of certain mental states extending on occasion (EMMT) would seem to promote a similar research methodology without the metaphysical commitments of EMT.

By contrast, if the radical enactivist approach is correct then this may well provoke a complete re-think of the methodology of cognitive science as well as traditional metaphysics about mind.

Issues for Enactivism

But we should be wary of hopping on board the enactivism train too soon. As indicated in the previous chapter, the enactivist approach is still in its infancy. And although it holds promise with regard to the formulation of an alternative mark of the mental that specifies a metaphysically extensive mind, there are a number of challenges that need to be met first.

Of primary concern is the Information-Processing Challenge. Because the principal explanatory resource availed of by enactivists is "the theory of self-organizing and autonomous dynamic systems" (Thompson 2007: 26) the worry is that it is in danger of undervaluing the role played by information processing mechanisms in making mental activity possible (Ramsey 2007, Clark 2008b). Unless it can meet this Information-Processing Challenge by providing a content-free account of how mentality emerges from engagement with the world then it is not clear that enactivism can provide a genuinely alternative approach to the question of the metaphysical extent of mind.

Furthermore, because the enactivist approach is in its infancy there is a requirement for further foundational work. In particular, appeals to self-organization in the form of autopoesis (following Varela et al. 1991) may entail that a living system and a cognitive system are co-extensive and that, therefore, the cognitive system does cannot extend beyond the boundaries of the organism (see Di Paolo 2009). Di Paolo admits that such a reading of autopoesis in the enactivist literature is possible but is

nevertheless adamant that "nothing like an internalist approach to mind is intended by the enactive approach" (2009: 12). That said, Di Paolo also recognizes that enactivism is far from the finished article and that a properly enactivist account of mind will take further work, particularly regarding appeals to autopoesis.

> It is a mistake to take the theory of autopoesis as originally formulated as a finished theory … autopoesis leaves many questions unanswered. In particular, several essential issues that could serve as a bridge between mind and life (like a proper grounding of teleology and agency) are given scant or null treatment in the primary literature. (Ibid: 12).

Despite these problems the radical enactivist approach offers more promise of developing a genuinely interesting EMT than current formulations. This is precisely because it rejects the representational view of mind and, in particular, it rejects non-derived content as a necessary condition on a state's being mental—the issue which has seen the current EMT debate as instigated by Clark and Chalmers hit stalemate.

Bibliography

Adams, F. & Aizawa, K. 2001. The Bounds of Cognition. *Philosophical Psychology* 14 (1): 14-64.

Adams, F. & Aizawa, K. 2005. Defending Non-Derived Content. *Philosophical Psychology* 18: 661-669.

Adams, F. & Aizawa, K. 2008. Why the Mind is Still in the Head: Challenges to Active Externalism. In Robbins, P. & Aydede, M. (eds.), *Cambridge Handbook of Situated Cognition*. Cambridge. Cambridge University Press: 78-95.

Adams, F. & Aizawa, K. Forthcoming. Defending the Bounds of Cognition. In Richard Menary (ed.), *The Extended Mind*. Ashgate.

Allen, C. 1992. Mental Content. *The British Journal for the Philosophy of Science* 43(4): 537-553.

Aizawa, K. Unpublished. Clark's Conditions on the Extended Mind are too Strong. Available at: http://personal.centenary.edu/~kaizawa/Clarks%20Conditions%0on%20Extended%20Mind%20are%20Too%20Strong.doc

Baier, K. 1964. The Place of a Pain. *The Philosophical Quarterly* 14 (55): 138-150.

Ballard, D. 1991. Animate Vision. *Artificial Intelligence* 48: 57-86.

Bartlett, G. 2008. Whither Internalism? How Internalists Should Respond to the Extended Mind Hypothesis. *Metaphilosophy* 39 (2): 163-184.

Bealer, G. 1994. Mental Properties. *The Journal of Philosophy* 91 (4): 185-208.

Bechtel, W. 1993. What Knowledge Must be in the Head in Order to Acquire Language. In Durham, W. & Velichovsky, B. (eds.) *Biological and Cultural Aspects of Language Development*. New Jersey: Lawrence Erlbaum.

Beckers et al. 1994. From local action to global tasks: Stigmergy and collective robotics. In Brooks, R. & Maes, P. (eds.) *Artificial Life 4*. MIT Press.

Beer, R. 1995a. A dynamical systems perspective on agent environment interaction. *Artificial Intelligence* 72: 173–215.

Beer, R. 1995b. Computational and dynamical languages for autonomous agents. In Port, R. & Van Gelder, T. (eds.) *Mind as Motion*. MIT Press.

Beer, R. & Gallagher, J. 1992. Evolving Dynamical Neural networks for adaptive behaviour. *Adaptive Behaviour* 1: 91–122.

Beer, R. & Chiel, H. 1993. Simulations of cockroach locomotion and escape. In Beer, R. et al. (eds.) *Biological Neural Networks in Invertebrate Neuroethology and Robotics*. Academic Press.

Blakemore, C. & Greenfield, S. (eds.) 1992. *Mindwaves*. Oxford: Basil Blackwell.

Block, N. 1980. Troubles with Functionalism. In Block, N. (ed.) *Readings in the Philosophy of Psychology, Vol. 1*. Cambridge: Harvard.

Block, N. 1996. What is Functionalism? *The Encyclopaedia of Philosophy Supplement*. New York: Simon & Schuster Macmillan.

Boden, M. 1988. *Computer models of mind*. Cambridge: Cambridge University Press.

Boden, M. (ed.) 1990. *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.

Braddon-Mitchell, D. & Jackson, F. 1996. *Philosophy of Mind and Cognition*. Oxford: Blackwell.

Branquinho J. 2001. *The Foundations of Cognitive Science*. Oxford: Oxford University Press

Bratman, M. et al. 1991. Plans and resource-bounded practical reasoning. In *Philosophy and AI: essays at the interface*.

Brook, A. 2006. My Blackberry and Me: Forever One or Just Friends? In Nyiri, K. (ed.), *Mobile understanding: The epistemology of ubiquitous communication*. Vienna. Passagen Verlag: 55-66.

Brooks, R. 1986. A Robust Layered Control System for a Mobile Robot. In *IEEE Journal of Robotics and Automation*, 14-23.

Brooks, R. 1991. Intelligence without Reason. In *Proceedings of the 12th International Joint conference on Artificial Intelligence*. Morgan Kauffman.

Brooks, R. 1993. A robot that walks: Emergent behaviours from a carefully evolved network. In Beer, R. et al. (eds.) *Biological Neural Networks in Invertebrate Neuroethology and Robotics*. Academic Press.

Brooks, R. 1994. Coherent Behaviour from Adaptive Processes. In Cliff, D. et al. (eds.) *From Animals to Animats 3*. MIT Press.

Brooks, R. 2002. *Robot: The Future of Flesh and Machines*. London: Penguin Books Ltd.

Burge, T. 1979. Individualism and the Mental. In P. French et al. (eds.), *Midwest Studies in Philosophy*, volume 4, Metaphysics. Minneapolis: University of Minnesota Press.

Chomsky, N. A. 1965. *Aspects of the theory of Syntax*. Cambridge. MA: MIT Press.

Churchland, P. M. 1988. *Matter and Consciousness* (revised edition). Cambridge, MA: Bradford/MIT Press.

Churchland, P. M. 1989. *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.

Churchland, P.M. 1995. *The Engine of Reason, the Seat of the Soul.* Cambridge, MA: MIT Press.

Churchland, P.S. 1986. *Neurophilosophy*. Cambridge, MA: Bradford/MIT Press.

Churchland, P.S. & Sejnowski, T. 1992. *The Computational Brain*. Cambridge, MA: Bradford/MIT Press.

Clark, A. 1989. *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Clark, A. 1997. *Being There: Putting Brain, Body and World Back Together Again*. Cambridge, MA: MIT Press.

Clark, A. 2001. Reasons, Robots and the Extended Mind. *Mind and Language* 16 (2): 121-145.

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, A. 2005. Intrinsic Content, Active Memory and the Extended Mind. *Analysis* 65 (1): 1-11.

Clark, A. 2007. Curing Cognitive Hiccups: A Defense of the Extended Mind. *Journal of Philosophy* 104 (4): 163-192.

Clark, A. 2008a. Pressing the Flesh: A Tension in the Study of the Embodied Embedded Mind? *Philosophy and Phenomenological Research* 76 (1): 37-59.

Clark, A. 2008b. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press. Oxford.

Clark, A. 2009. Spreading the Joy? Why the Machinery of Consciousness is (Probably) still in the head. *Mind* 118 (472): 963-993.

Clark, A. Forthcoming. Memento's Revenge: The Extended Mind, Extended. In Richard Menary (ed.), *The Extended Mind*. Ashgate.

Clark, A. & Chalmers, D. 1998. The Extended Mind. *Analysis* 58 (1): 7-19.

Clark, A. & Grush, R. 1999. Towards a Cognitive Robotics. *Adaptive Behaviour*, 7 (1): 5-16.

Coburn, R. C. 1966. Pains and Space. *The Journal of Philosophy*, 63 (13): 381-396.

Coleman, S. Forthcoming. There is no argument that the mind extends. Journal of Philosophy.

Copeland, J. (1993) *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.

Crane, T. 1998. Intentionality as the mark of the mental. In *Contemporary Issues in the Philosophy of Mind*. O'Hear, A. (ed.) Cambridge: Cambridge University Press: 229-251.

Cummins, R. 1989. *Meaning and Mental Representation*. Cambridge, MA: Bradford/MIT Press.

Dennett, D. 1978a. Why Not the Whole Iguana? *Behavioural and Brain Sciences*, 1: 103-4.

Dennett, D. 1978b. *Brainstorms*. Cambridge, MA: MIT Press.

Dennett, D. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dennett, D. & Kinsbourne, M. 1992. Time and the Observer: The Where and When of Consciousness in the Brain. *Behavioural and Brain Sciences* 15: 183-247.

Descartes, R. 1984-91. *The Philosophical Writings of Descartes* [3 volumes], J. Cottingham, R. Stoothoff, D. Murdoch & [volume 3 only] A. Kenny (trans.). Cambridge: Cambridge University Press.

Di Paolo, E. 2005. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4: 97–125

Di Paolo, E. 2009. Extended Life. Topoi 28: 9–21

Donald, M. 1991. *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Dretske, F. 1985. Machines and the Mental. *Proceedings and Addresses of the American Philosophical Association* 59 (1): 23-33.

Dretske, F. 1994. If you can't make one you don't know how it works. In *Midwest Studies in Philosophy XIX: Philosophical Naturalism*. University of Notre Dame Press.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dreyfus, H. L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Elder, C. L. 2001. Mental Causation versus Physical Causation: No Contest. *Philosophy and Phenomenological Research*, 62 (1): 111-127.

Fodor, J. 1975. *The Language of Thought*. New York: Crowell.

Fodor, J. A. 1980. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioural and Brain Sciences* 3: 63-72. Reprinted in *The Philosophy of Science*, R. Boyd, P. Gasper & J. D. Trout (eds.), MIT Press (1991): 651-669.

Fodor J. A. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Fodor J. A. 2009. Where is my mind? *London Review of Books*. 12 Feb 2009

Fodor, J. & Davies, M. 1986. Individualism and Supervenience. *Proceedings of the Aristotelian Society*, Supplementary Volumes, 60: 235-283.

Fodor, J. A. & Pylyshyn, Z. 1988. Connectionism and Cognitive Architecture: a Critical Analysis. *Cognition*, 28: 3-71.

Gallagher, S. 2008. Are Minimal Representations Still Representations? *International Journal of Philosophical Studies*, 16 (3): 351-369.

Gibson, J. J. 1950. *The Perception of the Visual World*. Boston: Houghton-Mifflin.

Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton-Mifflin.

Gibson, J. J. 1977. The Theory of Affordances. In Shaw, R & Bransford, J. eds. *Perceiving, Acting andKnowing: Towards an Ecological Psychology*. New Jersey: Lawrence Erlbaum.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Goldman, A. I. 1993. *Philosophical Applications of Cognitive Science*. Boulder, CO: Westview Press.

Grey, W. 1953. *The living brain*. Gerald Duckworth and Co. Ltd.

Grush, R. 2003. In defence of some "Cartesian" assumptions concerning the brain and its operation. *Biology and Philosophy* 18: 53-93.

Guha, R. V. & Lenat, D. B. 1990. 'CYC: A Midterm Report.' *AI Magazine*, 11(3): 33-59.

Guha, R. V. & Lenat, D. B. 1994. Enabling agents to work together. *Communications of the ACM*, 37(7): 127-142.

Guha, R.V. & Lenat, D.B. 1993. Response: Re:Cycling paper reviews. *Artificial Intelligence* 61:149-174.

Harnish, R. M. 2002. *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science*. Oxford: Blackwell.

Harvey, I. et al. 1994. 'Seeing the Light: Artificial evolution, real vision.' In Cliff, D. et al. eds. *From Animals to Animats 3*. Cambridge, MA: MIT Press.

Haugeland, J. 1981. Semantic engines: An introduction to mind design. In Haugeland, J. ed. *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press.

Holborow, L. C. 1966. Taylor on Pain Location. *The Philosophical Quarterly*, 16 (63): 151-158.

Holland, J. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Holly, W. J. 1986. The Spatial Coordinates of Pain. *The Philosophical Quarterly*, 36 (144): 343-356.

Hurley, S. 1998a. Vehicles, Contents, Conceptual Structure, and Externalism. *Analysis* 58 (1): 1-6.

Hurley, S. 1998b. *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Hurley, S. 2000. Consciousness in Action Clarifications. *Mind and Language* 15 (5): 556-561.

Hurley, S. 2001. Perception and Action: Alternative Views. *Synthese* 129: 3-40.

Hurley, S. 2003. Action, the Unity of Consciousness, and Vehicle Externalism. In Cleeremans, A. (ed.), *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.

Hurley, S. 2008. The Shared Circuits Model: How Control, Mirroring and Simulation Can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences* 31 (1): 1- 22.

Hurley, S. Forthcoming. Varieties of Externalism. In Richard Menary (ed.) *The Extended Mind*. Ashgate.

Hurley, S. & Noë, A. 2003. Neural Plasticity and Consciousness. *Biology and Philosophy* 18 (1): 131-168.

Hutchins, E. 1995. *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutto, D. D. 1999a. *The Presence of Mind.* Amsterdam: J Benjamins.

Hutto, D. D. 1999b. Cognition Without Representation? In A. Reigler & Markus F. Peschl (eds.), *Understanding Representation*. Plenum Press.

Hutto, D. D. 2005. Knowing What? Radical Versus Conservative Enactivism. *Phenomenology and the Cognitive Sciences* 4 (4):389-405

Hutto, D. D. 2008. Limited Engagements and Narrative Extensions. *International Journal of Philosophical Studies*, 16 (3): 419-444.

Hutto, D. D. 2009. Philosophical Clarification, its Possibility and Point. *Philosophia* 37 (4).

Hutto, D.D. Forthcoming. Presumptuous Naturalism: A Cautionary Tale. Available at: http://herts.academia.edu/documents/0073/0264/dh_Presumptous_Naturalism_-_15_Feb_10.pdf

Jackson, F. & P. Pettit. 1998. Functionalism and Broad Content. *Mind* 97 (387): 381-400.

Johnson, M. et al. (1994a) 'Evolving visual routines.' In Brooks, R. & Maes, P. eds. *Artificial Life 4*. Cambridge, MA: MIT Press.

Kim, J. 1996. *Philosophy of Mind*. Boulder, CO: Westview.

Kinsbourne, M. 1974. Mechanisms of Hemispheric Interaction in Man. In *Hemispheric Disconnection and Cerebral Function*, Marcel Kinsbourne and W. Lynn Smith, eds. Springfield, ill.;Charles C. Thomas: 239-59.

Kirsch, D. 1995. The intelligent use of space. *Artificial Intelligence*, 72: 1-52.

Kirsch, D. & Maglio, D. 1994. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18: 513-549.

Laird, J et al. 1987. SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33: 1-64.

Lenat, D.B. 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38 (11): 32-38.

Lenat, D. & Feigenbaum, E. 1992a. *Building Large Knowledge Based Systems: Representations and Inference in the CYC Project*. Addison-Wesley.

Lenat, D. & Feigenbaum, E. 1992b. On the Thresholds of Knowledge. In Kirsh, D. ed. *Foundations of Artificial Intelligence*. Cambridge, MA: MIT Press.

Lenat, D.B., Guha, R.V., Pittman, K., Pratt, D. & Shepherd, M. 1990. CYC: Toward Programs with common sense. *Communications of the ACM*, Aug 1990.

Lenat, D., Prakash, M. & Shepard, M. 1986. CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI Magazine*, 6 (4): 65-85.

Lenat, D. & Seely Brown, J. 1984. Why AM and EURISKO Appear to Work. *Artificial Intelligence*, 23: 269-294.

Lewis D. K. 1970. How to Define Theoretical Terms. *Journal of Philosophy* 67: 427-46.

Lewis D. K. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50: 249-58.

Luria, A. R., & Vygotsky, L. S. 1992. *Ape, Primitive Man and Child: Essays in the History of Behavior*. E. Rossiter, Trans. New York: Harvester Wheatsheaf.

Lycan, W. G. & J. J. Prinz (eds.). 2008. *Mind and Cognition: An Anthology, 3rd edition*. Wiley-Blackwell.

Maturana, H. 2002. Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybernetics and Human Knowing* 9:5–34

McGinn C. 1989. *Mental Content*. Oxford: Basil Blackwell.

Menary, R. 2006. Attacking the Bounds of Cognition. *Philosophical Psychology* 19 (3): 329–344

Menary, R. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan.

Menary, R. 2009. Intentionality, Cognitive Integration and the Continuity Thesis. *Topoi* 28: 31-43.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Millikan, R. G. 1991. Perceptual Content and Fregean Myth. *Mind* 100 (4): 439-59

Minsky, M. 1967. *Computation: Finite and Infinite Machines*. New Jersey: Prentice-Hall.

Minsky, M. 1968. Matter, mind, and models. In Minsky, M. ed. *Semantic information processing*. Cambridge, MA: MIT Press.

Minsky, M. 1974. A framework for representing knowledge. In *The Psychology of Computer Vision*. P.H. Wilson ed. McGraw-Hill, New York, 1975.

Minsky, M. 2000. Deep Issues: commonsense-based interfaces. *Communications of the ACM* 43,8: 66-73.

Moody, T. 1993. *Philosophy and Artificial Intelligence*. New Jersey: Prentice Hall.

Morris, K. J. 1984. In Defense of Methodological Solipsism: A Reply to Noonan. *Philosophical Studies* 45 (3): 399-411.

Myin, E., & Hutto, D. D. 2009. Enacting is enough. *Psyche* 15(1): 24-30.

Myin, E., O'Regan, K. 2009. Situated Perception and Sensation in Vision and Other Modalities: A Sensorimotor Approach. In *Cambridge Handbook of Situated Cognition*, (eds.) P. Robbins & M. Aydede: 185-200. New York: Cambridge University Press

Nelson, R. J. 1976. Mechanism, Functionalism, and the Identity Theory. *The Journal of Philosophy*, 73 (13): 365-385.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press.

Newell, A. & Simon, H. 1972. *Human Problem Solving*. New Jersey: Prentice-Hall.

Noë, A. 2003. Causation and Perception: the Puzzle Unravelled. *Analysis*, 63 (2): 93-100.

Noë, A. 2004. *Action in Perception*. MIT Press.

Nolfi, S. et al. 1994. How to evolve autonomous robots: different approaches in evolutionary robotics. In Brooks, R. & Maes, P. eds. *Artificial Life 4*. MIT Press.

Noonan, H. W. 1981. Methodological Solipsism. *Philosophical Studies* 40 (2): 269-274.

Noonan, H. W. 1985. Methodological Solipsism: A Reply to Morris. *Philosophical Studies* 48 (2): 285-290.

Norman, D. 1988. *The Psychology of everyday things*. Basic Books.

O'Regan, K. 1990. Eye Movements and reading. In Kowler, E. ed. *Eye Movements and their role in visual and cognitive processes*. Elsevier.

O'Regan, K. 1992. Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory.' *Canadian Journal of Psychology* 46 (3): 461-88.

O'Regan, K., Myin, E., Noë, A. 2005. Sensory Consciousness Explained (Better) in Terms of 'Corporality' and 'Alerting Capacity'. *Phenomenology and the Cognitive Sciences* 4:4 369 87.

O'Regan, K., Noë, A. 2001. A Sensorimotor Account of Vision and Visual Consciousness. *Behavioural and Brain Sciences* 24: 939 1031

O'Shaughnessy, B. 1957. The Location of Sound. *Mind*, 66 (264): 471 490.

Owens, J. 1993. Content, Causation, and Psychophysical Supervenience. *Philosophy of Science*, 60 (2): 242-261.

Owens, J. 1983. Functionalism and Propositional Attitudes. *Noûs*, 17 (4): 529-549.

Putnam, H. 1960. Minds and Machines. In S. Hook (ed.), *Dimensions of Mind: A Symposium*. New York University Press, New York.

Putnam, H. 1967a. The Mental Life of Some Machines. In Casteneda, H. N. ed. *Intentionality, Minds and Perception*. Detroit: Wayne State University Press.

Putnam, H. 1967b. Psychological Predicates. In W. Capitan & D. Merrill, (eds.), *Art, Mind and Religion*. Pittsburgh, PA: Pittsburgh University Press.

Putnam, H. 1975. The meaning of 'meaning'. In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.

Quine, W. V. 1985. States of Mind. *The Journal of Philosophy*, 82 (1): 5-8.

Quine, W.V. 1969. Natural Kinds. In *Ontological Relativity and Other Essays*. Columbia University Press, New York.

Ramsey, W. M. 2007. *Representation Reconsidered*. Cambridge: Cambridge University Press

Rapaport, W. J. 1995. 'Understanding Understanding: Syntactic Semantics and Computational Cognition.' *Philosophical Perspectives*, 9: 49-88.

Rowlands, M. 1989. Discussion of Jackson and Pettit, 'Functionalism and Broad Content'. *Mind*, New Series 98 (390): 269-275.

Rowlands, M. 1995a. Against Methodological Solipsism: the ecological approach. *Philosophical Psychology* 8 (10: 5-24.

Rowlands, M. 1995b. *Supervenience and Materialism*. Avebury. Aldershot.

Rowlands, M. 1997. Teleological Semantics**.** *Mind*, New Series, 106 (422): 279-303.

Rowlands, M. 1999. *The Body in Mind*. Cambridge: Cambridge University Press.

Rowlands, M. 2001. Consciousness and Higher-Order Thoughts. *Mind and Language* 16 (3): 290-310.

Rowlands, M. 2002. Two Dogmas of Consciousness. *Journal of Consciousness Studies* 9 (5-6): 158-80.

Rowlands, M. 2003a. Consciousness: The transcendentalist manifesto. *Phenomenology and the Cognitive Sciences* 2: 205-221.

Rowlands, M. 2003b. *Externalism: Putting Mind and World Back Together Again*. Acumen. Chesham.

Rowlands, M, 2006. *Body Language: Representation in Action*. Cambridge MA: Bradford Books/MIT Press.

Rowlands, M. 2009a. Extended Cognition and the Mark of the Cognitive. *Philosophical Psychology* 22(1): 1-19.

Rowlands, M. 2009b. Enactivism and the Extended Mind. *Topoi* 28 (1): 53-62.

Rorty, R. 1970. Incorrigibility as the Mark of the Mental. *The Journal of Philosophy*, 67 (12): 399-424.

Rumelhart, D. et al. 1986. *Parallel Distributed Processing: explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press.

Rupert, R. 2004. Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy* 101 (8): 389-428.

Ryle, G. 1949. *The Concept of Mind*. Re-print, 2000. Penguin Books, London.

Seager, W. 1983. Functionalism, Qualia and Causation. *Mind*, 92 (366): 174-188.

Searle, J. 1987. Minds and brains without programs. In Blakemore, C. & Greenfield, S. eds. *Mindwaves*. Oxford: Basil Blackwell.

Shapiro, L. A. 2008. Functionalism and Mental Boundaries. *Cognitive Systems Research* 9: 5-14.

Sikora, R. I. 1975. Rorty's New Mark of the Mental. *Analysis*, 35 (6): 192-194.

Sprevak, M. 2009. Extended Cognition and Functionalism. *Journal of Philosophy*, 106 (9): 503-527.

Stalnaker, R. 1989. On What's In the Head. *Philosophical Perspectives*, 3, Philosophy of Mind and Action Theory: 287-316.

Stalnaker, R. 1996. Varieties of Supervenience. *Noûs*, 30, Supplement: Philosophical Perspectives, 10, Metaphysics:221-241.

Sterelny, K. 1990. *The Representational Theory of Mind*. Oxford: Blackwell.

Sterelny, K. 2004. Externalism, Epistemic Artefacts and the Extended Mind. In Richard Schantz (ed.), *The Externalist Challenge*. De Gruyter.

Stich, S. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press

Sutton, J. 2006. Introduction: Memory, Embodied Cognition, and the Extended Mind. *Philosophical Psychology* 19(3): 281-289.

Taylor, J. G. 1962. *The Behavioural Basis of Perception*. New Haven, Conn.: Yale University Press.

Taylor, D. M. 1965. The Location of Pain. *The Philosophical Quarterly*, 15 (58): 53-62.

Taylor, D. M. 1966. The Location of Pain: A Reply to Mr. Holborow. *The Philosophical Quarterly*, 16 (65): 359-360.

Thompson, E. 2001. Empathy and consciousness. *The Journal of Consciousness Studies* 8:1–32

Thompson, E. 2007. *Mind in life: biology, phenomenology and the sciences of mind*. Cambridge: Harvard University Press.

Tye, M. 1992. Naturalism and the Mental. *Mind*, 101 (403): 421-441.

Van Gelder, T. 1995. What Might Cognition Be, If Not Computation? *The Journal of Philosophy*, 92 (7): 345-381.

Varela, F. 1997. Patterns of life: intertwining identity and cognition. *Brain and Cognition* 34:72–87.

Varela, F., Thompson, E., Rosch, E. 1991. *The embodied mind:cognitive science and human experience.* Cambridge, MA: MIT Press.

Wheeler, M. 1994. From activation to activity. *AISB Quarterly*, 87: 36 42.

Wheeler, M. 2005. Reconstructing the Cognitive World: The Next Step. MIT Press.

Wheeler, M. 2008a. Minimal Representing: A Response to Gallagher. *International Journal of Philosophical Studies* 16 (3): 371-376.

Wheeler, M. 2008b. Cognition in Context: Phenomenology, Situated Robotics and the Frame Problem. *International Journal of Philosophical Studies* 16 (3): 323-49.

Wilson, R. A. 1995. *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. Cambridge: Cambridge University Press.

Wilson, R. A. 2004. *Boundaries of the Mind: The Individual in the Fragile Sciences.* Cambridge: Cambridge University Press.

Wilson, R. A. Forthcoming. Meaning Making and the Mind of the Externalist. In: Menary, R. ed. *The Extended Mind*. Ashgate.